

基于 3 种建模法的食源性 ACE 抑制二肽 QSAR 研究

张艳萍,成 忠,俞远志

(浙江科技学院 生物与化学工程学院,杭州 310023)

摘 要: 以自组建的食源性血管紧张素转化酶(angiotensin I-converting enzyme, ACE)抑制二肽为研究样本,采用氨基酸描述子 VHSE(principal component score vector of hydrophilicity, steric, and electronic properties)对 ACE 抑制二肽进行表征后,比较偏最小二乘(partial least square regression, PLS)、支持向量机(support vector machine, SVM)及主成分分析(principal component analysis, PCA)-SVM 相结合的 3 种建模方法对 ACE 抑制二肽的 QSAR(quantitative structure-activity relationship)建模。结果显示,对于食源性 ACE 抑制二肽,3 个模型的拟合能力无明显差异,SVM 模型的预测能力略强;对其进行权重投影分析发现,C 末端氨基酸较 N 末端氨基酸对其活性的影响更为明显。

关键词: 血管紧张素转化酶;二肽;支持向量机;偏最小二乘;定量构效关系

中图分类号: R972

文献标志码: A

文章编号: 1671-8798(2015)03-0174-09

QSAR study of angiotensin I-converting enzyme inhibitory dipeptides based on three kinds of modeling methods

ZHANG Yanping, CHENG Zhong, YU Yuanzhi

(School of Biological and Chemical Engineering, Zhejiang University of
Science and Technology, Hangzhou 310023, China)

Abstract: A new ACE inhibitory dipeptides database was self-established. After the structures of dipeptides were characterized by using amino acid descriptors VHSE(principal component score vector of hydrophilicity, steric, and electronic properties), three kinds of modeling methods, namely partial least square regression(PLS), support vector machine(SVM), and principal component analysis(PCA) combined with SVM were used to establish the models of the QSAR of ACE inhibitory dipeptides, respectively. The results showed that there is no significant difference between the fitting abilities of the three models; the predictive abilities of

收稿日期: 2015-02-03

基金项目: 浙江省自然科学基金项目(LQ12C19004)

作者简介: 张艳萍(1976—),女,甘肃省瓜洲人,高级工程师,博士,主要从事功能肽的制备研究。

SVM model are stronger than other models. Moreover, the key structure factors relevant with dipeptide activities were studied. The results showed that the effect of the amino acid at C-terminal on ACE inhibitory activity of the dipeptides is more obvious than that at N-terminal.

Key words: angiotensin I-converting enzyme(ACE); dipeptide; support vector machine(SVM); partial least square regression(PLS); quantitative structure-activity relationship(QSAR)

血管紧张素转化酶(angiotensin converting enzyme, ACE)抑制肽是一类重要的功能活性肽,通过抑制体内 ACE 活性而起到显著的降血压作用^[1-3]。由于合成的 ACE 抑制剂易产生过敏、咳嗽等副作用,因此,从天然产物中分离制备具有显著 ACE 抑制活性的 ACE 抑制剂已经成为研究热点^[4]。而目前对于 ACE 抑制肽的研究,多是从混合基质中通过常规的生化分离手段来获得具有显著 ACE 抑制作用的活性肽。这种传统的研究模式存在工作效率低下、步骤繁琐等缺点,因此,快速、高效地发现具有显著活性的 ACE 抑制剂就显得尤为迫切。

ACE 抑制肽的定量构效关系(quantitative structure-activity relationship, QSAR)是在一系列已知的 ACE 抑制肽的结构及其 ACE 抑制活性之间建立数量依赖关系的数学模型,由此可以预测其他 ACE 抑制肽的活性,并为指导和设计新的 ACE 抑制因子提供理论依据^[5]。因此,深入研究食源性 ACE 抑制肽的 QSAR 模型对新的具有显著活性的 ACE 抑制肽的分子设计及类似降压药的研发具有重要意义。近年来,国内外研究者利用各种不同的描述子,通过不同的统计方法,进行了 ACE 抑制肽的 QSAR 研究^[6-9]。但仍然存在样本容量小、描述子不够全面与准确,以及建模方法不够科学等问题。本研究采用 VHSE(principal component score vector of hydrophilicity, steric, and electronic properties)描述子^[10],以文献报道的各种不同序列的食源性 ACE 抑制二肽作为样本,分别采用偏最小二乘(partial least squares, PLS)法、支持向量机(support vector machine, SVM)法^[11-12]及主成分分析^[13](principal component analysis, PCA)-支持向量机(SVM)相结合的统计方法,对收集的 ACE 抑制肽进行构效关系建模比较研究,建立了一个适用范围较广、回归能力显著、预测结果准确的 ACE 抑制肽的 QSAR 模型。

1 材料与方法

1.1 模型样本的收集

本研究广泛地收集了文献报道的食源性 ACE 抑制肽,共计 181 个二肽作为模型样本,以 ACE 抑制的 IC_{50} (μM)值作为活性指标。由于 IC_{50} 值相差倍数较大,为了提高分析结果的可靠性,本实验采用 $\log(IC_{50})$ 值作为其活性指标来进行数据的归一化处理。

1.2 VHSE 描述参数

VHSE 是基于 20 种天然氨基酸物理化学性质的一类描述子矢量。它是从 20 种天然氨基酸的 200 多种物理化学性质经主成分特征提取而产生的。并根据疏水性、立体和电性特征,将其再次进行主成分分析而得到的 8 个主成分。其中,VHSE1 和 VHSE2 代表氨基酸的疏水性特征;VHSE3 和 VHSE4 代表氨基酸的立体特征;VHSE5 到 VHSE8 则代表氨基酸的电性特征。20 种氨基酸的 VHSE 值见表 1。

1.3 氨基酸序列的数据转化

按照 ACE 抑制肽的氨基酸序列将其转化为一组多元数据。每个氨基酸由 8 个描述子来表示,即每个二肽的化学结构特征由 18 个描述子来表示。分别用 C_xV_y 表示对应的描述子,其中, x 表示肽段 C 端的位置, y 为描述子的序号。

表 1 天然氨基酸的 VHSE 描述参数

Table 1 VHSE scores for amino acids

氨基酸	缩写	VHSE1	VHSE2	VHSE3	VHSE4	VHSE5	VHSE6	VHSE7	VHSE8
Tyr	Y	0.61	1.60	1.17	0.73	0.53	0.25	-0.96	-0.52
Leu	L	1.36	0.07	0.26	-0.80	0.22	-1.37	0.08	-0.62
Phe	F	1.52	0.61	0.96	-0.16	0.25	0.28	-1.33	-0.20
Ala	A	0.15	-1.11	-1.35	-0.92	0.02	-0.91	0.36	-0.48
Lys	K	-1.17	0.70	0.70	0.80	1.64	0.67	1.63	0.13
Arg	R	-1.47	1.45	1.24	1.27	1.55	1.47	1.30	0.83
Gly	G	-0.20	-1.53	-2.63	2.28	-0.53	-1.18	2.01	-1.34
Glu	E	-1.18	0.40	0.10	0.36	-2.16	-0.17	0.91	0.02
Thr	T	-0.34	-0.51	-0.55	-1.06	-0.06	-0.01	-0.79	0.39
Val	V	0.76	-0.92	-0.17	-1.91	0.22	-1.40	-0.24	-0.03
Ile	I	1.27	-0.14	0.30	-1.80	0.30	-1.61	-0.16	-0.13
Asp	D	-1.15	0.67	-0.41	-0.01	-2.68	1.31	0.03	0.56
Asn	N	-0.99	0.00	-0.37	0.69	-0.55	0.85	0.73	-0.80
Trp	W	1.50	2.06	1.79	0.75	0.75	-0.13	-1.01	-0.85
Pro	P	0.22	-0.17	-0.50	0.05	-0.01	-1.34	-0.19	3.56
Ser	S	-0.67	-0.86	-1.07	-0.41	-0.32	0.27	-0.64	0.11
Gln	Q	-0.96	0.12	0.18	0.16	0.09	0.42	-0.20	-0.41
Met	M	1.01	-0.53	0.43	0	0.23	0.10	-0.86	-0.68
His	H	-0.43	-0.25	0.37	0.19	0.51	1.28	0.93	0.65
Cys	C	0.18	-1.67	-0.46	-0.21	0	1.20	-1.61	-0.19

1.4 QSAR 建模

采用 SYSTAT12 软件,以抑制活性值为因变量,VHSE 描述子为基础的结构表征参数为自变量,采用 PLS 法、SVM 法、PCA-SVM 相结合的方法分别对二肽的多元数据进行回归建模。模型的拟合能力采用相关系数 R^2 和均方根误差 RMSEE 检验, R^2 越大 RMSEE 越小,说明模型的拟合能力越好;由于建立 QSAR 模型的最终目的在于预测未知化合物的活性,因此,模型的预测能力较拟合能力更为重要。留一法交互验证是目前较为广泛使用的一种验证方法,即依次从样本中抽出 1 个样本,用余下的 $n-1$ 个样本建立模型,并预测该样本的活性,依次循环,直到每个样本都被抽出并预测过,最后计算预测残差平方和(predictive residual sum of squares, PRESS)及交互验证 R^2 ,以 PRESS 最小为原则确定模型当中所需要的主成分个数。

一个具有准确预测能力的模型应该具备以下条件: $R^2 > 0.6$ 且 $0.85 < k < 1.15$ 。其中, R^2 为预测集样本实验值对预测值回归方程的决定系数; R_0^2 和 R'^2_0 分别为预测集样本实验值与预测值通过原点回归方程的决定系数(R_0^2 为预测值—实验值; R'^2_0 为实验值—预测值); k 和 k' 分别为相应回归方程的斜率。同时,采用经典的预测均方根(root mean square error of prediction, RMSEP)来评价所建模型的外部预测能力。

1.5 模型的泛化性能检验

将贻贝中分离得到的 ACE 抑制二肽 Val-Trp (VW)进行模型的泛化性能检验,用以评价模型的预测能力。

2 结果与讨论

2.1 食源性 ACE 抑制二肽的 QSAR 研究

自组建的食源性 ACE 抑制二肽库共包括 181 个二肽样本,采用 VHSE 描述子对肽库中的每个二肽进行定量描述后,随机选取 163 个样本作为训练集,另外 18 个样本作为测试集用于模型的外部验证。分别应用 PLS、SVM 及 PCA-SVM 3 种建模方法对该体系进行构效关系研究。其中,PLS 与 PCA-SVM 方法模型中所需的特征变量(成分)个数是由交叉验证试验方式选定,且各初始变量在特征变量中的贡献可由权重向量解释。具体为:二肽的 PLS 模型需提取 1 个特征成分,它概括解释预测自变量的能力达

24.20%,概括解释性质活性变量的能力则可达63.22%;而PCA-SVM模型则需提取9个特征成分,它们对预测自变量的概括解释能力达98%。由上述3种建模方法构建的各QSAR模型对ACE抑制活性的测定值和预测值见表2。

表2 ACE抑制二肽的氨基酸序列及其生物活性值

Table 2 ACE inhibitor sequences with observed and calculated activities for dipeptides

样本 编号	肽序列	测定值	预测值			样本 编号	肽序列	测定值	预测值		
			PLS	SVM	PCA-SVM				PLS	SVM	PCA-SVM
1	YP	2.86	2.62	2.61	2.64	45	YY	1.53	1.62	1.59	1.34
2	LW	1.24	1.31	1.24	1.07	46	DG	1.09	3.90	2.67	2.23
3	FP	2.50	2.49	2.46	2.57	47	WL	1.81	2.09	2.07	1.85
4	AP	1.46	2.85	2.80	2.32	48	SY	1.82	1.86	2.06	1.33
5	KP	1.34	2.89	2.21	1.76	49	GY	2.41	2.48	2.87	2.25
6	RP	1.32	2.85	2.29	1.81	50	FY	0.57	1.49	1.42	1.10
7	GP	2.56	3.48	3.19	2.72	51	NY	1.51	2.07	2.15	1.43
8	EP	3.08	3.05	3.04	2.87	52	NF	1.67	2.14	2.37	1.73
9	TP	2.46	2.68	2.80	2.52	53	WA	2.44	2.72	2.39	2.28
10	VP	1.70	2.53	2.54	2.17	54	WM	1.98	1.98	2.08	1.90
11	GI	3.11	3.09	3.12	3.11	55	MW	1.00	1.37	1.19	1.07
12	DF	2.56	2.08	2.48	2.13	56	MY	2.29	1.66	1.73	1.25
13	YQ	2.80	2.34	2.25	1.92	57	YL	1.91	2.21	2.25	1.85
14	IW	0.30	1.17	0.98	0.71	58	HY	1.42	1.83	1.86	1.28
15	IY	0.57	1.46	1.58	1.11	59	VY	0.85	1.53	1.69	1.18
16	NP	3.36	3.07	3.02	2.84	60	RF	1.97	1.92	2.32	1.93
17	DM	2.78	2.49	2.68	2.43	61	GG	3.94	4.38	3.54	3.53
18	DL	3.30	2.60	2.92	2.70	62	GR	2.41	2.89	2.94	2.97
19	VW	0.15	1.24	1.01	0.70	63	PR	0.61	2.02	1.92	1.60
20	YW	1.02	1.33	1.24	1.29	64	RY	1.02	1.85	1.96	1.40
21	VY	1.34	1.53	1.69	1.18	65	FY	1.15	1.49	1.42	1.10
22	WN	1.71	2.50	2.05	2.05	66	AW	0.98	1.56	1.40	0.99
23	CF	0.30	1.81	1.90	1.53	67	VW	0.04	1.24	1.01	0.70
24	EY	0.43	2.05	2.09	1.46	68	GW	1.64	2.20	2.25	1.88
25	MF	1.65	1.73	1.99	1.68	69	DW	1.11	1.72	2.05	1.62
26	FE	0.16	2.66	2.03	1.78	70	IP	2.11	2.46	2.46	2.16
27	AY	2.00	1.85	2.14	1.42	71	IR	2.92	1.86	2.44	2.38
28	DY	2.00	2.01	2.25	1.72	72	QK	2.95	2.50	2.74	2.71
29	RY	1.71	1.85	1.96	1.40	73	RL	3.39	2.45	2.72	2.63
30	GY	1.86	2.48	2.87	2.25	74	RW	1.20	1.56	1.81	1.23
31	IY	0.32	1.46	1.58	1.11	75	VK	1.11	2.19	2.48	2.17
32	LY	1.59	1.60	1.80	1.32	76	VQ	3.11	2.25	2.63	2.56
33	KY	1.11	1.89	1.73	1.23	77	WI	1.91	2.10	2.23	2.06
34	FY	1.40	1.49	1.42	1.10	78	YG	3.04	3.51	3.20	3.32
35	VY	1.20	1.53	1.69	1.18	79	YP	2.95	2.62	2.61	2.64
36	AF	1.18	1.91	2.42	1.94	80	LY	0.83	1.60	1.80	1.32
37	DF	1.67	2.08	2.48	2.13	81	KW	0.20	1.60	1.41	1.03
38	LF	2.10	1.67	2.33	1.94	82	LY	0.81	1.60	1.80	1.32
39	KF	2.06	1.96	2.14	1.83	83	YL	1.21	2.21	2.25	1.85
40	SF	2.11	1.93	2.22	1.74	84	AY	1.28	1.85	2.14	1.42
41	TF	1.25	1.75	2.11	1.69	85	IY	0.38	1.46	1.58	1.11
42	VF	0.96	1.60	2.12	1.79	86	AW	1.08	1.56	1.40	0.99
43	IF	2.97	1.52	2.13	1.80	87	LW	0.83	1.31	1.24	1.07
44	GF	2.44	2.55	3.01	2.67	88	LW	1.70	1.31	1.24	1.07

续表 2

样本 编号	肽序列	测定值	预测值			样本 编号	肽序列	测定值	预测值		
			PLS	SVM	PCA-SVM				PLS	SVM	PCA-SVM
89	LF	2.54	1.67	2.33	1.94	136	KF	1.45	1.96	2.14	1.23
90	YL	2.09	2.21	2.25	1.85	137	FY	0.57	1.49	1.42	1.83
91	GP	3.08	3.48	3.19	2.64	138	VW	1.03	1.24	1.01	1.10
92	VP	2.76	2.53	2.54	2.72	139	VY	1.64	1.53	1.69	0.70
93	GF	2.85	2.55	3.01	2.17	140	IW	1.09	1.17	0.98	1.18
94	VW	0.20	1.24	1.01	2.67	141	VY	1.05	1.53	1.69	0.71
95	FQ	1.71	2.21	2.11	0.70	142	IA	2.18	2.67	2.56	1.18
96	VY	1.41	1.53	1.69	1.92	143	WL	1.48	2.09	2.07	2.42
97	TF	1.95	1.75	2.11	1.18	144	VW	0.40	1.24	1.01	1.85
98	LY	1.51	1.60	1.80	1.69	145	IW	0.67	1.17	0.98	0.70
99	AF	1.88	1.91	2.42	1.32	146	RP	1.96	2.85	2.29	0.71
100	IY	1.02	1.46	1.58	1.94	147	FY	0.81	1.49	1.42	1.81
101	YV	2.76	2.47	2.51	1.11	148	GT	3.76	3.33	3.45	1.10
102	YE	2.80	2.79	2.19	2.41	149	GE	3.85	3.65	2.89	3.50
103	GW	1.48	2.20	2.25	2.05	150	GD	3.96	3.51	2.69	3.30
104	GY	2.32	2.48	2.87	1.88	151	VG	3.04	3.42	3.33	3.25
105	GP	2.65	3.48	3.19	2.25	152	IG	3.08	3.35	3.25	3.10
106	GF	2.80	2.55	3.01	2.72	153	RG	3.08	3.74	2.91	3.15
107	GM	3.15	2.96	3.38	2.67	154	YG	3.30	3.51	3.2	3.04
108	GA	3.30	3.70	3.45	3.14	155	AG	3.40	3.74	3.56	3.32
109	GL	3.40	3.08	3.24	3.41	156	KG	3.51	3.78	3.14	3.50
110	GH	3.49	3.25	3.29	3.01	157	FG	3.57	3.38	3.10	3.15
111	GR	3.51	2.89	2.94	3.36	158	MG	3.68	3.55	3.31	3.30
112	GS	3.58	3.54	3.50	2.97	159	WG	3.77	3.39	3.04	3.56
113	GV	3.66	3.34	3.18	3.56	160	HG	3.80	3.72	3.27	3.15
114	GK	3.73	3.14	3.07	3.21	161	EG	3.87	3.94	3.07	3.47
115	GQ	3.73	3.21	3.57	3.14	162	SG	3.93	3.75	3.43	3.02
116	LF	3.52	1.67	2.33	3.46	163	LG	3.94	3.49	3.34	3.48
117	KP	1.71	2.89	2.21	1.94	164	TG	4.00	3.57	3.42	3.45
118	FL	1.20	2.08	2.03	1.76	165	QG	4.00	3.74	3.46	3.35
119	VY	1.25	1.53	1.69	1.79	166	DG	4.15	3.90	2.67	3.65
120	IL	1.74	2.05	2.18	1.18	167	PG	4.23	3.51	2.94	2.23
121	VY	1.55	1.53	1.69	2.01	168	VW	0.20	1.24	1.01	2.60
122	IY	0.79	1.46	1.58	1.18	169	VF	1.72	1.60	2.12	0.70
123	AW	1.27	1.56	1.40	1.11	170	AY	1.94	1.85	2.14	1.79
124	FY	1.63	1.49	1.42	0.99	171	RP	2.26	2.85	2.29	1.42
125	VW	0.52	1.24	1.01	1.10	172	AF	2.28	1.91	2.42	1.81
126	IW	0.18	1.17	0.98	0.70	173	AP	2.36	2.85	2.80	1.94
127	LW	1.37	1.31	1.24	0.71	174	RF	2.36	1.92	2.32	2.32
128	FY	0.22	1.49	1.42	1.07	175	VP	2.62	2.53	2.54	1.93
129	VY	1.76	1.53	1.69	1.10	176	AP	2.43	2.85	2.80	2.17
130	GQ	3.75	3.21	3.57	1.18	177	IY	0.43	1.46	1.58	2.32
131	TP	3.32	2.68	2.80	3.46	178	VW	0.23	1.24	1.01	1.11
132	TK	3.21	2.34	2.60	2.52	179	MW	0.58	1.37	1.19	0.70
133	YH	0.71	2.39	2.06	2.44	180	RW	1.34	1.56	1.81	1.07
134	KW	1.03	1.60	1.41	1.91	181	KP	1.48	2.89	2.21	1.23
135	KY	0.89	1.89	1.73	1.03						

对3个模型的回归结果作图分析,得到如图1和图2所示的结果。图1和图2分别为3个模型的计算值与实验值的回归图及实验值与预测值通过原点的回归图。从图中可观察到:实验值与拟合值及实验值与预测值之间的线性关系均良好。

3个模型的内部和外部验证结果(表3)表明:3个模型的 R^2 、 R_{ext}^2 、 R_0^2 和 $R_0'^2$ 均大于0.60,其中,PCA-SVM模型具有最强的拟合能力,表现在其具有最高的 R^2 值及最低的RMSEE,但其 R_{ext}^2 、 R_0^2 、 $R_0'^2$ 值则无明显优势。并且,其RMSEP值较PLS模型和SVM模型为大, k 及 k' 亦不接近1,因此,就预测能力而言,PLS及SVM模型较PCA-SVM模型略强,而PLS和SVM模型的预测能力则基本相当。

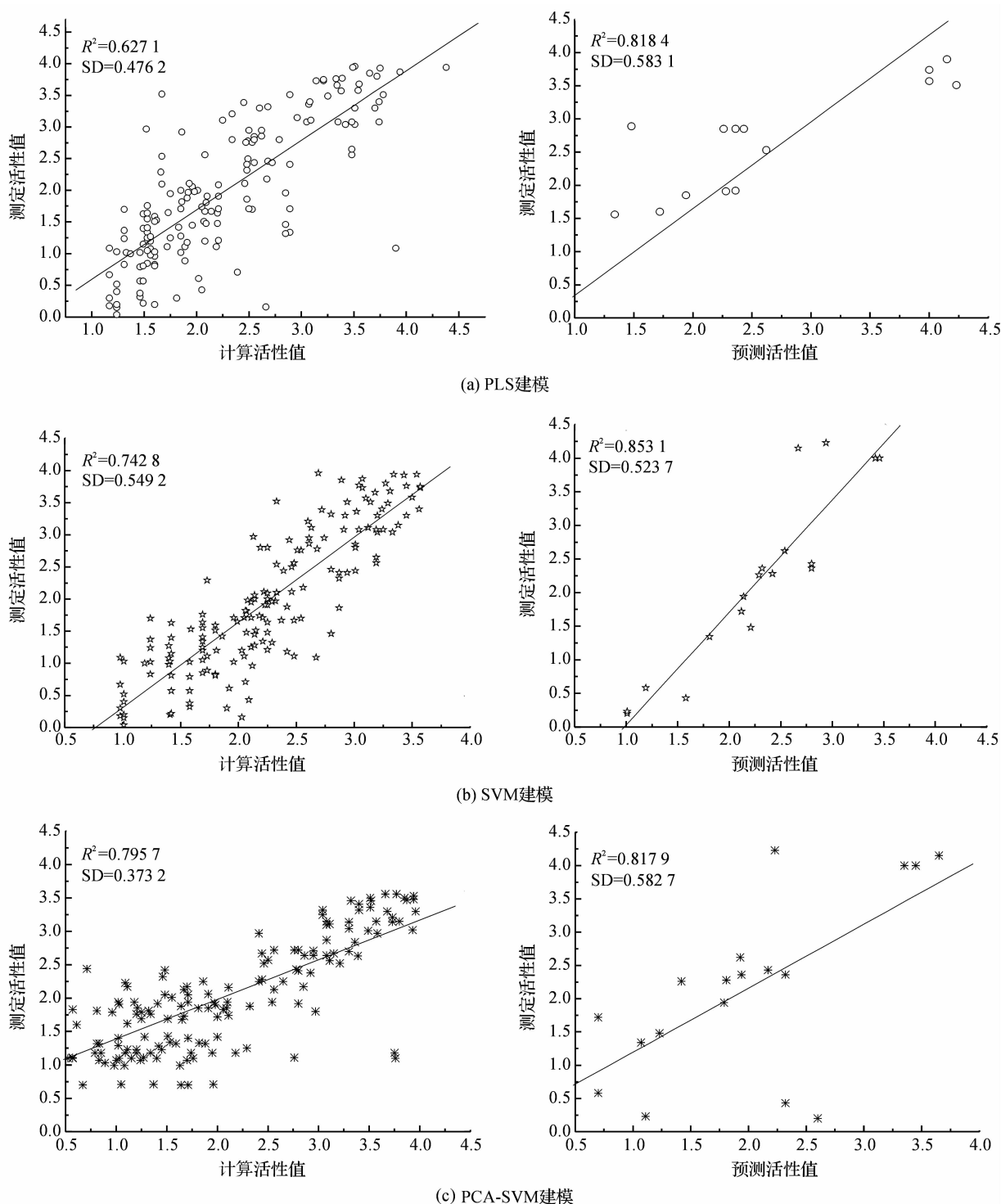


图1 ACE抑制二肽的计算值与实验值回归图

Fig. 1 Regression between calculated and observed activities of ACE inhibitory dipeptides

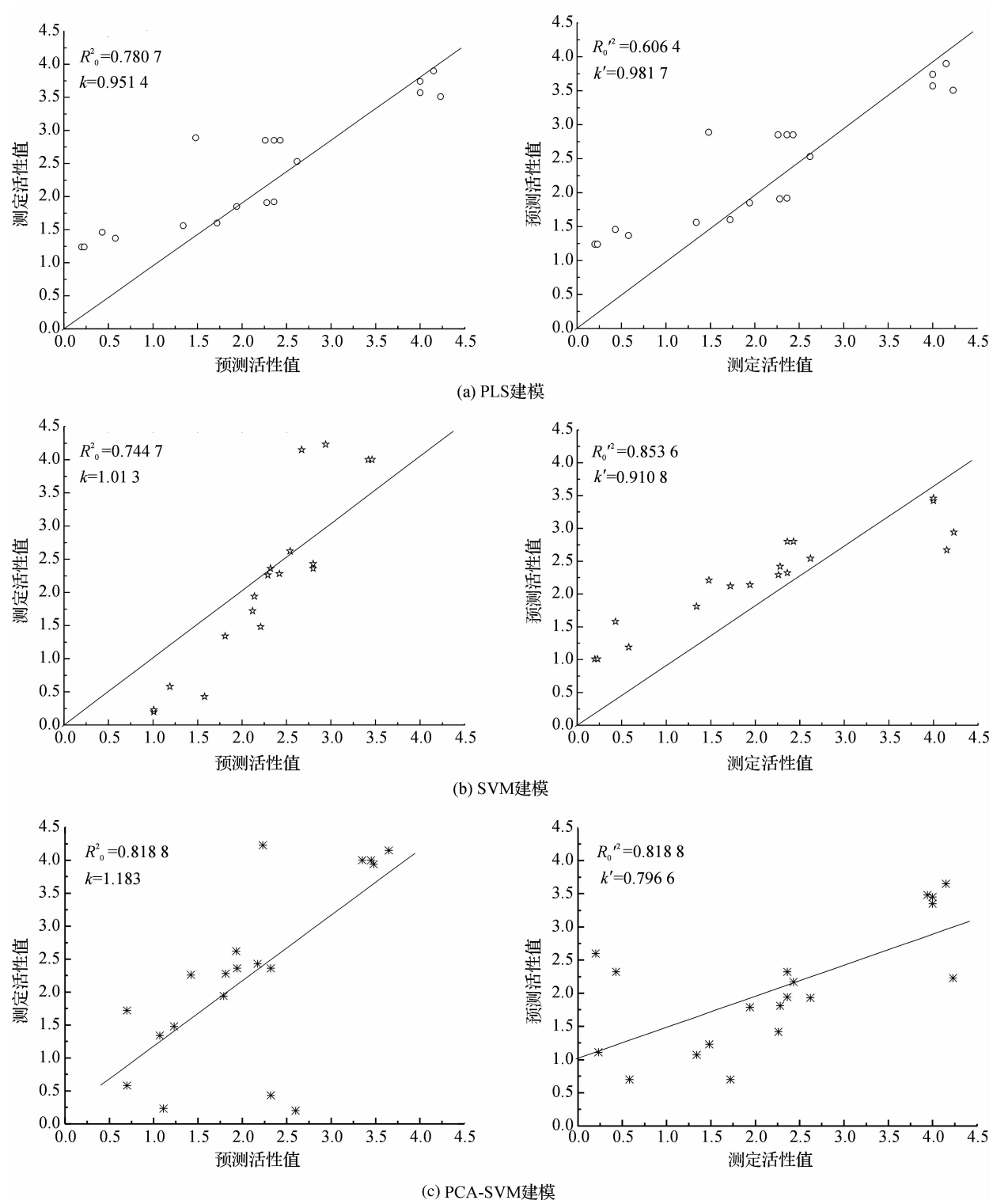


图 2 ACE 抑制二肽的预测值与实验值通过原点回归图

Fig. 2 Regression through origin between predicted and observed activities ACE inhibitory dipeptides

表 3 ACE 抑制二肽的验证结果

Table 3 Results of internal and external validation of model for ACE inhibitory dipeptides

模型	R^2	RMSEE	R^2_{ext}	RMSEP	R^2_0	R'^2_0	k	k'
PLS	0.627 1	0.718 3	0.818 4	0.656 9	0.780 7	0.606 4	0.951 4	0.981 7
SVM	0.742 8	0.698 1	0.853 1	0.654 6	0.744 7	0.853 6	1.013	0.910 8
PCA-SVM	0.795 7	0.494 8	0.817 9	0.712 0	0.818 8	0.818 8	1.183	0.796 6

从图3中可以看出各描述子变量与活性 Y 之间的相关性。比较图3(a)和(b)发现,采用不同的模型构建方法,所得结果具有一定的相似性:C端氨基酸与其ACE抑制活性呈正相关;N端氨基酸的V4则呈显著负相关。图3(a)的具体表现为:对于该二肽体系,C1V1、C1V2、C1V3、C1V5、C1V6、C2V1、C2V2、C2V3、C2V5、C2V8与其ACE抑制活性呈现正相关;C1V4、C1V7、C1V8、C2V4、C2V6、C2V7则与其呈负关联。并且,二肽的C端氨基酸较N端氨基酸对其活性的影响更为明显。其中,C端氨基酸各因素对二肽的ACE抑制活性的影响依次为V3(立体性质) $>$ V2(疏水性质) $>$ V7(电性性质),而N端氨基酸各因素对活性的影响依次为V4(立体性质) $>$ V7(电性性质) $>$ V1(疏水性质)。具体地,在2个氨基酸位置,较高的V3、V2、V1和较低的V7均有利于其活性,对照各氨基酸的参数发现:二肽中含有Trp、Tyr和Phe有利于其与ACE的竞争结合;N端出现具有较低V4的Val、Ile、Thr时有利于其与ACE的结合,上述结论与文献[6]报道基本一致。

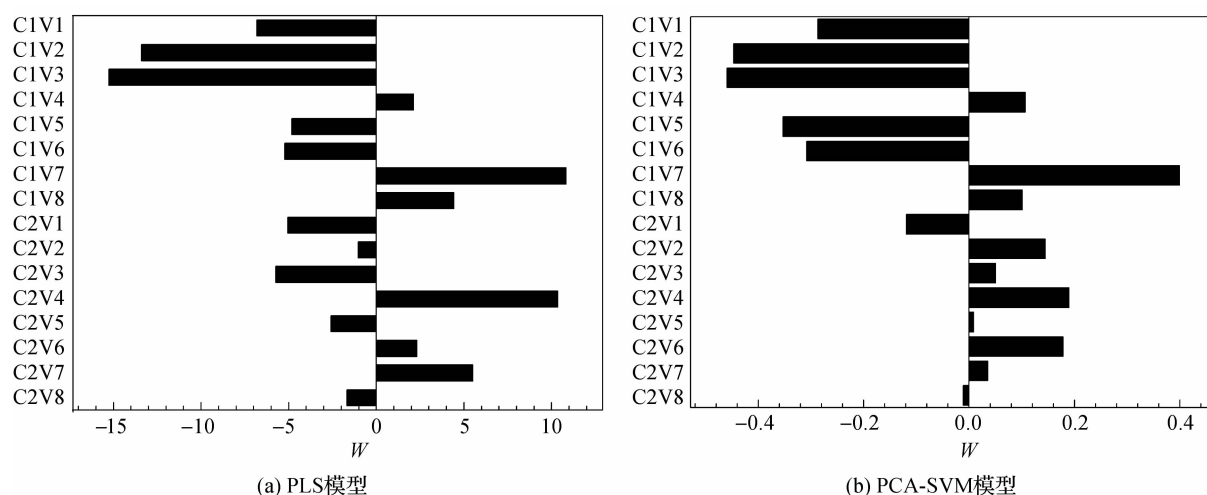


图3 ACE抑制二肽的变量投影重要性指示图

Fig. 3 VIP plot for ACE inhibitory dipeptides

2.2 贻贝中的ACE抑制二肽对QSAR模型的检验

笔者从贻贝蛋白的酶解物中分离得到了1个具有显著ACE抑制作用的二肽,其氨基酸序列为VW,并测定得到其ACE抑制的 IC_{50} 值分别为 $1.7 \mu M^{[14]}$ 。通过建立食源性ACE抑制二肽的PLS模型、SVM模型及PCA-SVM模型,各模型对于贻贝中已知序列及活性的ACE抑制肽的活性预测结果如表4所示。

表4 贻贝中ACE抑制肽的活性预测结果

Table 4 Prediction and experimental validation of ACE inhibitory peptides from MEPI hydrolysates

ACE抑制肽	实验值 (log IC_{50})	PLS模型		SVM模型		PCA-SVM模型	
		预测值	偏差	预测值	偏差	预测值	偏差
VW	0.23	1.24	1.01	1.01	0.78	0.70	0.47

由表4可知,3个模型对于贻贝蛋白酶解物中的ACE抑制肽的活性预测结果有较大差异。对VW而言,PLS模型和SVM模型的预测结果偏差较大,而PCA-SVM模型的预测结果相对较准确。因此,以贻贝中分离得到的ACE抑制肽的活性测定值和预测值进行模型的泛化性能检验可知,本研究已建立的3个ACE抑制肽的QSAR模型能够对已知结构的ACE抑制二肽的活性进行预测,其中,SVM模型与PCA-SVM模型的预测结果较为准确,即SVM模型及PCA-SVM模型具有较PLS模型更强的预测能力。预测结果出现的较大偏差可能与检验模型的样本数量偏小有关。

3 结论

1)在运用VSHE描述子进行氨基酸结构表征的基础上,比较了PLS、SVM及PCA-SVM对食源性

ACE 抑制二肽的建模结果。结果表明:对于 ACE 抑制二肽,3 个模型的拟合能力无明显差异,而 SVM 模型的预测能力略强。

2)对于 ACE 抑制二肽,C 末端氨基酸对二肽的 ACE 抑制活性的影响较 N 末端氨基酸更为明显;对照各氨基酸的参数发现:二肽中含有 Trp、Tyr 和 Phe 有利于其与 ACE 的竞争结合;N 端出现具有较低 V4 的 Val、Ile、Thr 时有利于其与 ACE 的结合。

3)采用贻贝中分离得到的 ACE 抑制二肽对 3 个模型进行泛化性能检验,结果表明:本研究建立的 3 个 ACE 抑制肽的 QSAR 模型能够对已知结构的 ACE 抑制肽的活性进行预测,其中 SVM 模型及 PCA-SVM 模型具有较 PLS 模型更强的预测能力。

参考文献:

- [1] 刘静,管晓,彭剑秋. 神经网络在 ACE 抑制肽 QSAR 中的应用研究[J]. 分析测试学报,2012,31(10):1260-1265.
- [2] Jamdar S N, Rajalakshmi V, Pednekar M D, et al. Influence of degree of hydrolysis on functional properties, antioxidant activity and ACE inhibitory activity of peanut protein hydrolysate[J]. Food Chemistry,2010,121(1):178-184.
- [3] Qu W J, Ma H L, Pan Z L, et al. Preparation and antihypertensive activity of peptides from porphyra yezoensis[J]. Food Chemistry,2010,123(1):14-20.
- [4] Ko S C, Kang M C, Lee J K, et al. Effect of angiotensin I-converting enzyme (ACE) inhibitory peptide purified from enzymatic hydrolysates of styela plicata[J]. European Food and Technology,2011,233: 915-922.
- [5] 梅虎. 肽的定量构效关系研究[D]. 重庆:重庆大学,2005.
- [6] Wu J, Aluko R E, Nakai S. Structural requirements of angiotensin I converting enzyme inhibitory peptides: Quantitative structure activity relationship study of di-and tri peptides [J]. Journal of Agricultural and Food Chemistry,2006,54(3):732-736.
- [7] Wu J, Aluko R E, Nakai S. Structural requirements of angiotensin I converting enzyme inhibitory peptides: Quantitative structure-activity relationship modeling of peptides containing 4-10 amino acid residues[J]. QSAR and Combinatorial Science,2006,25(10):873-880.
- [8] Pripp A H, Isaksson T, Stepaniak L, et al. Quantitative structure-activity relationship modeling of ACE-inhibitory peptides derived from milk proteins[J]. European Food Research and Technology,2004,219(6):579-583.
- [9] Pripp A H. Docking and virtual screening of ACE inhibitory dipeptides[J]. European Food Research and Technology,2007,225(4):589-592.
- [10] Mei H, Liao Z H, Zhou Y, et al. A new set of amino acid descriptors and its application in peptide QSARs[J]. Biopolymers,2005,80(6):775-786.
- [11] 周鹏,曾晖,李波,等. 支持向量机分类和回归用于肽的 QSAR 研究[J]. 化学通报,2006,69(5):342-346.
- [12] 印家健,李梦龙,文志宁,等. 支持向量回归用于氨基酸描述符在肽 QSAR 建模中的性能评价[J]. 四川大学学报,2006,43(2):396-402.
- [13] 梅虎,周原,孙立力,等. 一种新的氨基酸描述子及其在肽 QSAR 中的应用[J]. 物理化学学报,2004,20(8):821-825.
- [14] 张艳萍. 贻贝蛋白中 ACE 抑制肽的制备及其构效关系研究[D]. 杭州:浙江工商大学,2010.