J Mol Model (2011) 17:1599–1606
DOI 10.1007/s00894-010-0862-x

ORIGINAL PAPER

QSAR study on angiotensin-converting enzyme inhibitor oligopeptides based on a novel set of sequence information descriptors

Xiaoyu Wang · Juan Wang · Yong Lin · Yuan Ding · Yuanqiang Wang · Xiaoming Cheng · Zhihua Lin

Received: 13 July 2010 / Accepted: 27 September 2010 / Published online: 13 October 2010

Abstract A novel set of descriptors G-scale was derived from 457 physicochemical properties of the natural amino acids. The descriptors were then applied to study on quantitative structure-activity relationships (QSARs) of nine peptide datasets of angiotensin-converting enzyme inhibitor (ACE-inhibitor) oligopeptides (between dipeptides and decapeptides) by using partial least square (PLS) regression. The multiple correlation coefficients (R^2) and leave one out cross validation values (Q^2) of PLS models are better than or close to the results of references. The results show that the descriptors proposed here may be a useful structural expression method, and they may be hopefully used in biological activity study of ACE-inhibitor oligopeptides.

Keywords Amino acids descriptors · Angiotensin-converting enzyme inhibitor oligopeptides · Partial least square regression · Quantitative structure-activity relationship

Electronic supplementary material The online version of this article (doi:10.1007/s00894-010-0862-x) contains supplementary material, which is available to authorized users.

X. Wang · J. Wang · Y. Lin · Y. Ding · Y. Wang · Z. Lin (✉)
College of Pharmacy and Bioengineering,
Chongqing University of Technology,
Chongqing 400054, People's Republic of China

X. Cheng (✉)
Institute of Respiratory Diseases, Xinqiao Hospital,
Third Military Medical University,
Chongqing 400037, People's Republic of China
e-mail: cxm65420021@yahoo.com

Z. Lin
e-mail: zhlin@cqut.edu.cn

Introduction

Angiotensin-converting enzyme (ACE) plays an important role in maintaining human blood dynamic balanced in renin-angiotensin system (RAS). ACE catalyzes inactive Angiotensin Idecapeptides to generate strongly vasoconstrictive Angiotensin IIoctapeptides [1], and the latter is a kind of strong vasoconstrictor and adrenal cortex of aldosterone freed activation agent. Therefore, ACE has become an appropriate target to hypotensive drugs [2]. ACE-inhibitors can inhibit the generation of angiotensin II and control hypertension [3]. It has become a prodrug to treat hypertension, heart disease, diabetes mellitus and nephrosis etc., so to develop new efficient ACE-inhibitor drugs has more pressing practical significance.

With the exploitation and development of new drug, quantitative structure-activity relationships (QSARs) has been brought into the spotlight in drug protophase research by using mathematical models to quantitatively describe the relationship between target molecular structures and biological activities [4]. If the relationship between the structures of ACE-inhibitor peptides and biological activities can be confirmed, the blood pressure dropping drugs will be successfully synthesized on this basis [5].

In the 1960s, Sneath et al. [6] first expressed peptide sequences by using semi-quantitative experimental parameters of 20 coded amino acids and successfully predicted the activities of hypophamine. Kidera et al. [7] collected 188 properties of the 20 natural amino acids and employed factor analysis to obtain 10 orthogonal factors which were the most important for determining the three dimensional structures of proteins. Hellberge et al. [8–10] used principal component analysis (PCA) for 29 physicochemical properties to each of 20 natural amino acids,

including electrostatic, stereo, hydrophobic properties, and was called '3-Z' scale. Subsequently, two new variables were extracted from '3-Z' scale residue matrixes by Sandberg et al. [11], together with the three properties above, which was called '5-Z' scale. Based on the three dimensional descriptors, isotropic surface area (ISA) and electronic charge index (ECI), Collantes et al. [12] established good 3D-QSAR models. In recent years, a series of descriptors were proposed, which can well represent the structural characteristics of the amino acids for QSAR models, such as MS-WHIM [13], MARCH-INSIDE [14], VHSE [15], T-scales [16], VSW [17], V [18], HESH [19, 20] and ST-scale [21] etc.

Almost all the descriptors mentioned above were derived from the principal component analysis (PCA) of the data matrixes, which may cursorily explained physicochemical properties of the amino acids, and may need the definite information of amino acids, such as hydrophobicity, molecular volume, net charge etc. Therefore, a novel amino acid descriptors G-scale was acquired and optimized based on previous work [18], and was then employed to represent the structures of ACE-inhibitory peptides, QSAR modelings of ACE-inhibitor oligopeptides were constructed by using partial least square (PLS) regression method. The results show that the descriptors may be a useful structural expression method for the study on QSAR of ACE-inhibitor peptides.

Methods

Partial least square regression

Partial least square (PLS) regression is a widely used modeling method. PLS regression combines basic functions of regression model, principal component analysis (PCA) and canonical analysis. An excellent model should have both favorable estimated abilities for any internal samples and outstanding predictive abilities for any external samples [22, 23]. In this study, PLS regression was completed by SIMCA-P 10.0 software.

Variable selections and structural characterizations

The novel set of amino acid descriptors including eight kinds of parameters were derived from 457 kinds of physicochemical properties of the amino acid index database, which was classified into three sorts of parameters including hydrophobic, steric and electric properties (Supporting information Table S1). Each sort of properties were used to characterize ACE-inhibitor dipeptide set based on QSAR model with stepwise multiple regression (SMR), and the optimal parameters of amino acid were applied to characterize peptide sequences based on the cross validation value of SMR model. So eight appropriate parameters were collected as the descriptors for characterize ACE-inhibitor peptide

Table 1 Descriptors of 20 natural occurring amino acids

AA	G1	G2	G3	G4	G5	G6	G7	G8
Ala (A)	0.870	1.520	0.400	1.800	19.200	0.058	5.080	0.007
Cys(C)	1.520	1.520	0.500	-16.500	40.470	0.128	2.950	-0.037
Asp(D)	0.660	1.520	0.800	5.050	40.470	0.081	5.960	-0.024
Glu(E)	0.670	1.520	1.300	12.000	51.790	0.064	6.040	0.007
Phe(F)	2.870	1.520	0.700	-34.500	79.910	0.065	4.360	0.038
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
Lys(K)	1.640	1.520	0.400	14.600	65.850	0.095	4.930	0.018
Leu(L)	2.170	1.520	0.600	-11.000	57.620	0.070	8.030	0.052
Met(M)	1.670	1.520	0.300	-10.000	64.140	0.055	2.610	0.003
Asn(N)	0.090	1.520	0.900	-5.600	42.180	0.091	5.750	0.005
Pro(P)	2.770	1.520	0.900	-86.200	43.560	0.068	4.840	0.240
Gln(Q)	0.000	1.520	0.700	6.300	52.130	0.098	4.240	0.049
Arg(R)	0.850	1.520	0.300	12.500	74.770	0.085	4.750	0.044
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
Val(V)	1.870	1.900	0.400	5.630	43.900	0.053	6.070	0.057
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

Table 2 The sequences and ACE-inhibitory activities of 58 dipeptides

NO.	peptide	obsd	calcd	NO.	peptide	obsd	calcd
1	VW	5.80	5.30	30	KG	2.49	2.57
2	IW	5.70	5.55	31	FG	2.43	2.42
3	IY	5.43	4.81	32	GS	2.42	2.21
4	AW	5.00	4.94	33	GV	2.34	2.38
5	RW	4.80	5.07	34	MG	2.32	2.54
6	VY	4.66	4.56	35	GK	2.27	2.83
7	GW	4.52	4.39	36	GE	2.27	2.42
8	VF	4.28	4.11	37	GT	2.24	2.26
9	AY	4.06	4.20	38	WG	2.23	2.43
10	IP	3.89	4.07	39	HG	2.20	2.06
11	RP	3.74	3.58	40	GQ	2.15	2.56
12	AF	3.72	3.76	41	GG	2.14	1.83
13	GY	3.68	3.65	42	QG	2.13	2.18
14	AP	3.64	3.46	43	SG	2.07	2.11
15	RF	3.64	3.88	44	LG	2.06	2.42
16	VP	3.38	3.81	45	GD	2.04	2.40
17	GP	3.35	2.91	46	TG	2.00	2.16
18	GF	3.20	3.21	47	EG	2.00	2.21
19	IF	3.03	4.37	48	DG	1.85	2.26
20	VG	2.96	2.73	49	PG	1.77	1.87
21	IG	2.92	2.99	50	LA	3.51	2.74
22	GI	2.92	2.88	51	KA	3.42	2.90
23	GM	2.85	2.88	52	RA	3.34	2.83
24	GA	2.70	2.16	53	YA	3.34	2.65
25	YG	2.70	2.32	54	AA	3.21	2.70
26	GL	2.60	2.45	55	FR	3.04	3.23
27	AG	2.60	2.38	56	HL	2.49	2.68
28	GH	2.51	2.94	57	DA	2.42	2.58
29	GR	2.49	2.64	58	EA	2.00	2.54

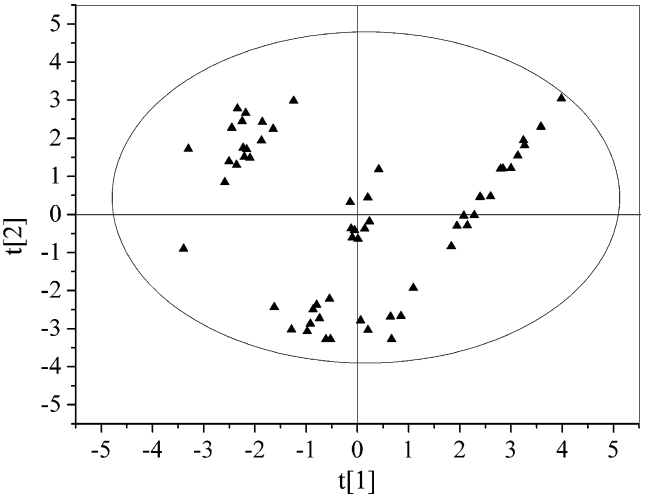

Fig. 2 Plots of PLS scores of ACE-inhibitor dipeptides

sequences. The collection of parameters was finished by Matlab 7.0 software. The eight parameters include hydrophobicity, STERIMOL minimum width of the side chain, loss of side chain hydropathy by helix formation, optical rotation, side chain molecular volume, frequency of the 4th residue in turn, AA composition of EXT of multi-spanning proteins and net charge index [24], and they were encoded as G1~G8 (Table 1). For a set of peptide analogues, the chemical structures would be now characterized by describing each varied amino acid position with 8 G-scale values. For example, the chemical structures of dipeptides would be described by 16 (8×2) variables. Thus, a set of peptide sequences varied in n positions can be described by 8×n variables.

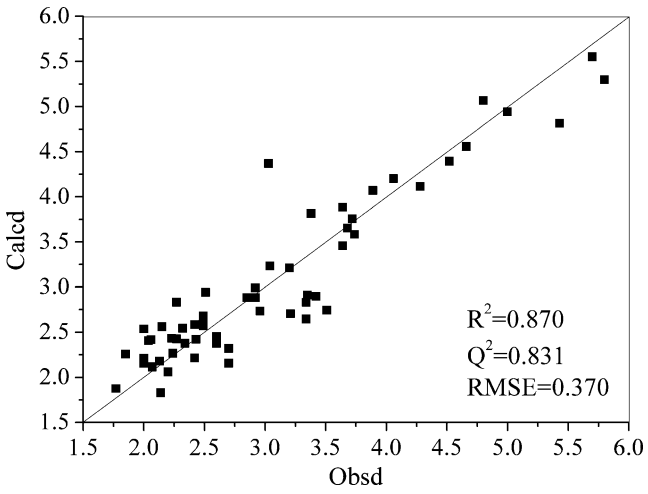

Fig. 1 Plots of calculated and observed activities of ACE-inhibitor dipeptides

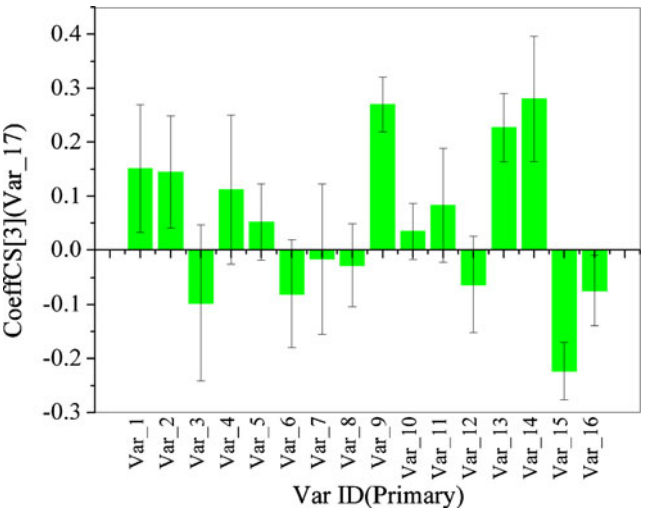

Fig. 3 Summary of PLS regression coefficients

Results and discussion

QSAR analysis on ACE-inhibitor dipeptides

As a classical dataset for QSAR study, a set of 58 dipeptides of ACE-inhibitors was used to test the effectiveness of amino acid descriptors [25]. The structural information of each dipeptide was quantified by 16 G-scale variables. Then the QSAR modeling was performed by using PLS regression method. It yielded a squared correlation coefficient (R^2) of 0.870, and after 'leave-one-out' cross validation, it yielded a correlation coefficient (Q^2) of 0.831. The calculated biological activity data of the 58 ACE-inhibitor dipeptides are shown in Table 2.

The results of QSAR modeling are illustrated in Fig. 1, where the relationship between the calculated and the observed values is presented. It shows that almost all the samples are uniformly dispersed around the 45 degrees line except the 19#, which shows the calculated and observed values are interrelated well. Plots of PLS scores of ACE-inhibitor dipeptides are shown in Fig. 2. All the samples locate in Hotelling T^2 confidence ellipse with 95% confidence. PLS regression coefficients are summarized in Fig. 3. Looking at the coefficients values, it is evident that variables of amino acid properties are important for the bioactivities of ACE-inhibitor diepetides. For the first position, G1, G2, G4 and G5 have a positive correlation with biological activities, while G3, G6, G7 and G8 have a negative correlation, and G7, G8 tend to 0 related to biological activity values. For the second position, G1, G2, G3, G5 and G6 are positive to biological activities, while G4, G7 and G8 are negative. For the two positions, G1, G2 and G5 are positively tended to biological activities, while G7 and G8 are negative to them. The second position is more relevant to biological activities than the first, and G1, G5, G6 and G7 are the most important property parameters to ACE-inhibitor dipeptides. For the ACE-inhibitor dipeptides, amino acid residues with information of hydrophobic and stereo characteristics are most important to biological activities. For both residues, amino acid residue with large bulk chain and hydrophobic side chains are preferred and Gly is the most favorable amino acid via statistics. It may obtain high biological activities of ACE-inhibitor dipeptides by the alteration of these amino acid residues.

Table 3 The comparison between QSAR models of ACE-inhibitor dipeptides

N0.	Descriptors	Model	Q^2	R^2	RMSE
1	z-scales [10]	PLS	nd	0.770	nd
2	GRID [26]	PLS	nd	0.744	0.50
3	ISA-ECI [12]	PLS	nd	0.700	nd
4	VSTV [27]	PLS	0.767	0.789	0.46
5	SSIA-AM1 [28]	PLS	0.699	0.769	0.49
6	SSIA-PM3 [28]	PLS	0.773	0.789	0.47
7	SZOTT [29]	PLS	0.753	0.878	0.33
8	T-scales [16]	PLS	0.786	0.845	0.39
9	VSW [17]	PLS	0.784	0.868	0.37
10	G-scales	PLS	0.831	0.870	0.37

The results of this scale descriptor model and other descriptor models are given in Table 3. It indicates that G-scale model is superior to most of the other traditional descriptor models. The value of R^2 of this scale model is slightly lower than that of SZOTT model, while its predictability of Q^2 is obviously superior to other models.

QSAR analysis on ACE-inhibitor tripeptides

The 55 ACE-inhibitor tripeptides were collected from literature [25]. An optimal model including 24 variables

Table 4 The sequences and ACE-inhibitory activities of 55 tripeptides

NO.	peptide	obsd	calcd	NO.	peptide	obsd	calcd
1	VVV	1.63	1.69	29	RRR	1.77	1.86
2	RPG	3.09	3.16	30	PPP	1.86	1.90
3	GRP	0.48	0.51	31	FFF	1.20	1.25
4	LLL	1.35	1.35	32	RGP	1.73	1.69
5	LGG	2.49	2.47	33	PGR	2.67	2.69
6	GLG	2.45	2.45	34	GGV	1.99	2.16
7	GGL	1.63	1.60	35	GVV	1.82	1.73
8	LLG	2.33	2.34	36	PPG	3.18	3.25
9	LGL	1.52	1.49	37	PGG	3.14	3.12
10	GLL	1.47	1.47	38	PGP	1.82	1.77
11	FGG	2.79	2.81	39	GPG	2.65	2.72
12	GFG	2.53	2.53	40	GGP	1.28	1.25
13	GGF	1.11	1.09	41	PGI	2.23	2.22
14	FFG	2.71	2.75	42	KPK	2.63	2.24
15	FGF	1.29	1.31	43	ADA	2.17	2.23
16	GFF	1.02	1.03	44	LDL	1.42	1.38
17	GGG	2.61	2.59	45	GEG	2.28	2.28
18	YGG	3.07	3.06	46	LEL	1.19	1.17
19	GYG	2.33	2.28	47	RGP	1.73	1.69
20	GGY	1.35	1.39	48	PIP	1.69	1.48
21	YYG	2.79	2.75	49	FPF	1.32	1.44
22	YGY	1.82	1.85	50	VIF	0.78	0.77
23	GYY	1.07	1.08	51	KPF	1.51	1.39
24	YYY	1.54	1.55	52	RPF	1.59	1.66
25	FIV	2.04	2.09	53	PPF	1.68	1.75
26	FPP	1.50	1.59	54	VYP	0.82	0.91
27	FPK	2.45	2.29	55	YPF	1.60	1.69
28	PFP	1.74	1.71				

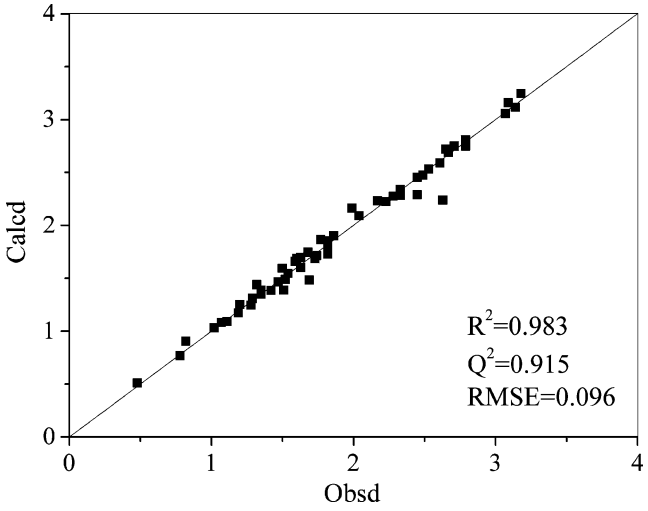

Fig. 4 (**a**) PLS coefficient plots of ACE-inhibitor dipeptides (**b**) Plots of calculated and observed activities of ACE-inhibitor tripeptides

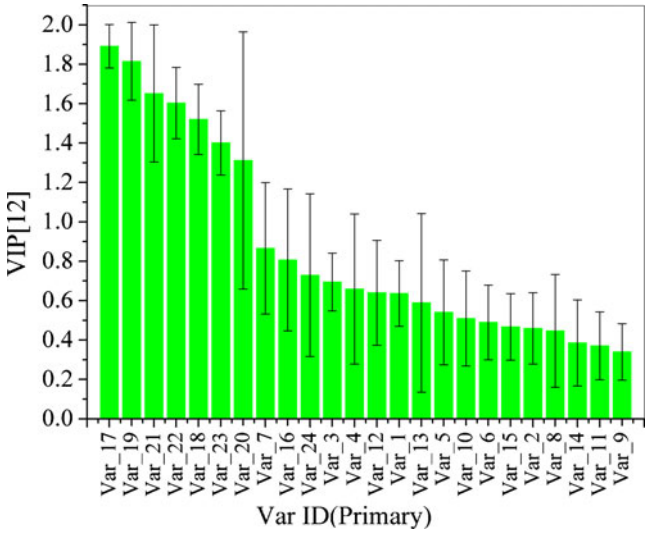

Fig. 6 VIP plots for ACE-inhibitor tripeptides by PLS

was constructed by PLS regression method, it yielded a squared correlation coefficient (R^2) of 0.983, and a root mean square error (RMSE) of 0.096, which reflects the relativities of the errors between the experimental values and the estimated values. After 'leave-one-out' cross validation, it yielded a correlation coefficient (Q^2) of 0.915. The sequences and ACE-inhibitory activity datas of 55 tripeptides are shown in Table 4.

The plots of calculated and observed activities of ACE-inhibitor tripeptides are shown in Fig. 4. All the samples uniformly dispersed around the 45 degrees line, which shows the calculated and observed values are interrelated well. From the plots of Distance to PLS model in X space of ACE-inhibitor tripeptides (Fig. 5), it can be seen that almost all the samples except 43#, 45# and 50# distances in X space are less than 1.528, which is a significant test

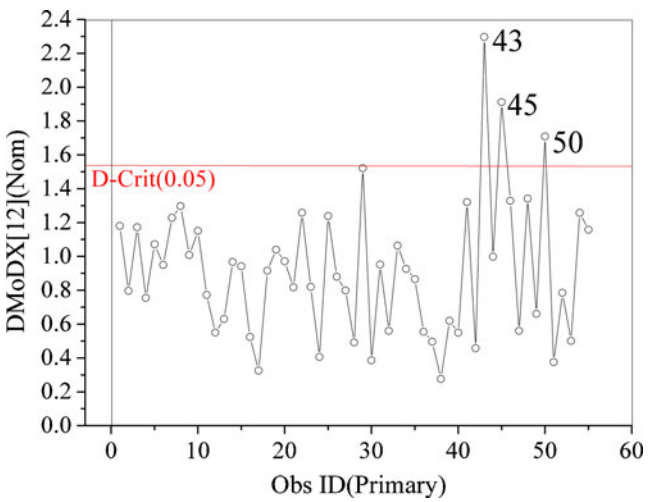

Fig. 5 Plots of distance to PLS model in X space of ACE-inhibitor tripeptides

critical value. It shows that the characterization effectiveness for three samples is worse the than average of other samples. VIP plots of the PLS models are summarized in Fig. 6. It is very obvious that the third position is most influential to biological activities, and is followed by the first position, then the second position. The most influential property parameters to ACE-inhibitor tripeptides are G1, G5 and G4 and Gly is the most favorable amino acid via statistics. According to the influence of biological activities level, eight parameters characterized in the third position were arranged in a proper order as G1, G3, G5, G6, G2, G7, G4 and G8. The least influential factors are G1, G3 and G6 in the second position. For the first position, eight parameters were arranged in a proper order as G7, G3, G4, G1, G5, G6, G2 and G8. For the second position, the order is G8, G4, G5, G2, G7, G6, G3 and G1. Hydrophobic and stereo properties are more relevant to the biological activities of ACE-inhibitor tripeptides. Especially, hydrophobic is the most influential parameter.

The comparison between QSAR models of ACE-inhibitor tripeptides is shown in Table 5. It indicates that this scale descriptor model is more superior to 3-z scale model, and the value of Q^2 of G-scale model is slightly lower than that of V scale, while the value of R^2 and RMSE are superior to V scale. So G-scale can be considered as optimal descriptors for QSAR.

Table 5 The comparison between QSAR models of ACE-inhibitor tripeptides

N0.	Descriptors	Model	Q^2	R^2	RMSE
1	3-z Scales [25]	PLS	0.426	0.500	0.404
2	V [18]	MLR	0.943	0.970	0.154
3	G-scales	PLS	0.915	0.983	0.096

Table 6 The sequences and ACE-inhibitory activities of nonapeptides

NO.	peptide	obsd	calcd	NO.	peptide	obsd	calcd
1	IVGRPRHQG	0.79	0.66	11	RPKHPIKHQ	1.13	1.09
2	IVGRPRHQG	0.38	0.66	12	YPFPGPIPN	1.17	1.22
3	QVSLNSGYY	1.36	1.48	13	LVYPFPGPI	2.22	2.28
4	LDAQSAPLR	2.80	2.79	14	VIGSPPQIN	3.00	3.08
5	AMKPWIQPK	2.78	2.55	15	RMLGQTPTK	1.53	1.42
6	ALKAWSVAR	0.48	0.55	16	LTQTPVVPP	2.20	2.34
7	GPAGAPGAA	1.57	1.50	17	VVYPWTQRF	1.01	0.90
8	EWPRPQIPP	-0.25	-0.32	18	YYAPFDGIL	1.92	1.98
9	YANPAVVRP	0.89	0.77	19	AFKAWAVAR	0.23	0.33
10	ALNEINQFY	2.34	2.26				

QSAR analysis on ACE-inhibitor nonapeptides

The 19 nonapeptides were collected from literature [30]. An optimal model including eight variables (V57, V20, V30, V27, V6, V13, V5, and V67) was constructed by SMR- PLS method. It yielded 4 principal components, which explained 61.69%, 25.87%, 9.08% and 1.67% of Y variable variance respectively, and the first two principal components explained vast majority of the variance. It yielded a squared correlation coefficient (R^2) of 0.983, a root mean square error (RMSE) of 0.138. After 'leave-one-out' cross validation, it yielded a correlation coefficient (Q^2) of 0.894. The sequences and ACE-inhibitory activity datas of nonapeptides are shown in Table 6, and the comparison between QSAR models of ACE-inhibitor nonapeptides is shown in Table 7. It can be seen that this scale descriptor model is superior to z-scale models.

The plots of calculated and observed activities of ACE-inhibitor nonapeptides are shown in Fig. 7. It shows calculated and observed values are interrelated well. From the plots of PLS scores of ACE-inhibitor nonapeptides (Fig. 8), it can be found that all the samples are located in Hotelling T^2 confidence ellipse with 95% confidence. The PLS loading plots of ACE-inhibitor nonapeptides are shown in Fig. 9. It shows each variable contribution to biological activities. In the first principal component, V57, V20, V30, V27 and V67 have a positive correlation with biological activities, and V57 (hydrophobicity) is most relevant to them; V6, V13 and V5 are negative with biological activities, V13 (side chain molecular volume) is most relevant to them. In the second principal component, V57, V20 and V30 are positive with biological activities, and V30 is most relevant to them. V27, V5, V13, V6 and V67 are negative with biological activities, and V67 is most relevant to them. V57 and V30 are most influential to biological activities respectively in two principal components, and V30 is more influential than V57 for ACE-inhibitor nonapeptides. V6 tend to 0 related to biological activity values in two principal components. The first and the forth position are the main positions for ACE-inhibitor nonapeptides, and Pro is the most favorable amino acid via statistics. G6 is the most influential parameter to biological activities, followed is G1. The main property parameters affect biological activities of ACE-inhibitor nonapeptides are hydrophobic and stereo properties. Especially, frequency of the 4th residue characterized in the fourth position is a main factor to the biological activities of ACE-inhibitor nonapeptides.

Table 7 The comparison between QSAR models of ACE-inhibitor nonapeptides

N0.	Descriptors	Model	Q^2	R^2	RMSEE
1	3-z Scales [30]	PLS	0.234	0.908	nd
2	5-z Scales [30]	PLS	0.255	0.689	nd
3	G-scales	PLS	0.894	0.983	0.138

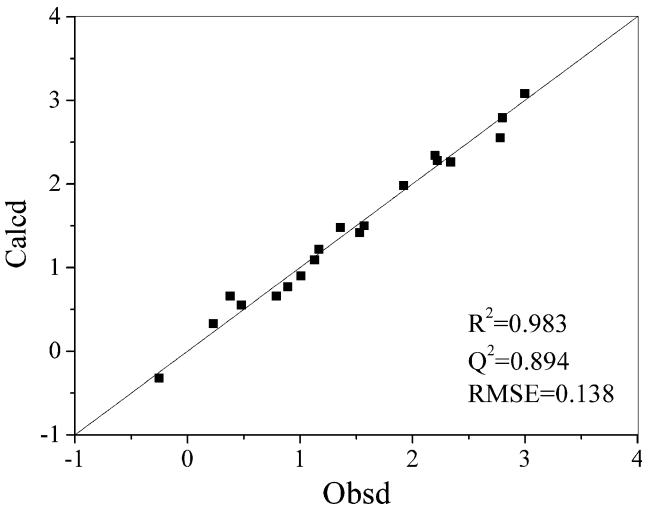

Fig. 7 Plots of calculated and observed activities of ACE-inhibitor nonapeptides

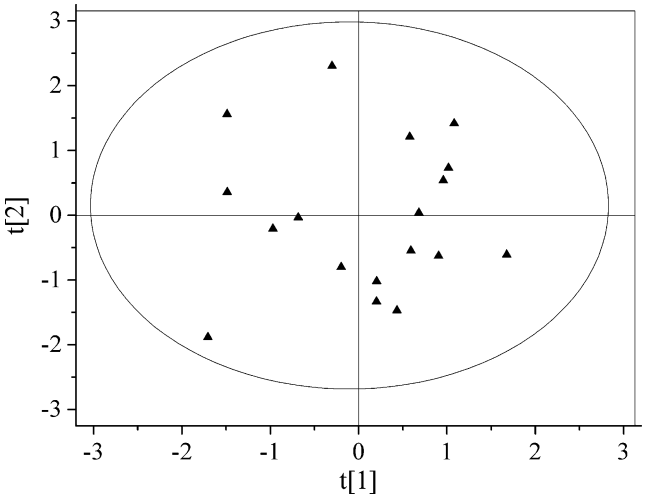

Fig. 8 Plots of PLS scores of ACE-inhibitor nonapeptides

QSAR analysis on ACE-inhibitor between tetrapeptides and decapeptides

The amino acid sequences and observed activity datas between tetrapeptides and decapeptides were collected from literature [30], the calculated and observed biological activities list in the supporting information Table S2. All the optimal models were constructed by SMR-PLS method respectively. Parameters of QSAR studies between tetrapeptides and decapeptides are shown in Table 8. The results of G-scale model and z-scale models are shown in Table 9. Through the research, for penta-, hepta- and deca- peptides, stereo property is the most influential parameters. For hexa- and octa- peptides, hydrophobic and stereo properties are both the preferential parameters, and for tetrapeptides, stereo and electrical characteristics are the most influential parameters.

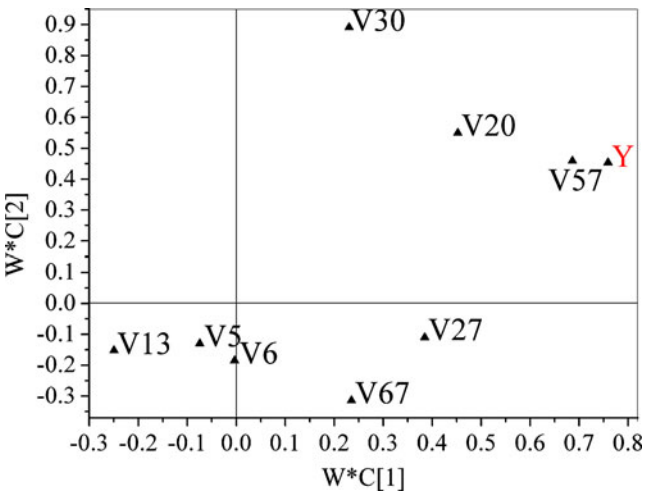

Fig. 9 PLS loading plots of ACE-inhibitor nonapeptides

Table 8 QSAR model results of ACE-inhibitor between tetrapeptides and decapeptides

Peptide	N	Variables	Q^2	R^2	Comps
Tetra-	79	V16, V2, V27, V30, V13, V18	0.294	0.403	1
Penta-	73	V13, V39, V17, V25	0.325	0.382	1
Hexa-	61	V25, V47, V2, V14, V1, V3, V28, V22	0.267	0.555	2
Hepta-	31	V22, V25, V14, V44, V9	0.417	0.709	2
Octa-	29	V17, V42, V43, V8	0.577	0.639	1
Deca-	13	V7, V37, V80, V73	0.529	0.905	3

Conclusions

In this study, a set of amino acid descriptors G-scale including hydrophobic, stereo and electrical characteristics was derived from 457 physicochemical parameters of amino acid index data after gradual screening and optimization. G-scale was then systematically applied to the study of quantitative structure-activity relationships (QSARs) on NINE peptide datasets of angiotensin-converting enzyme inhibitor oligopeptides (between dipeptides and decapeptides) by using partial least square (PLS) regression. This paper provided a simple and effective method for QSAR study on ACE-inhibitor

Table 9 The comparisons between QSAR models of ACE-inhibitor between tetrapeptides and decapeptides

Peptide	N0.	Descriptors	Model	Q^2	R^2
Tetra-	1	3-z Scales [30]	PLS	0.235	0.398
	2	5-z Scales [30]	PLS	0.309	0.485
	3	G-scales	PLS	0.294	0.403
Penta-	1	3-z Scales [30]	PLS	0.14	0.587
	2	5-z Scales [30]	PLS	0.365	0.523
	3	G-scales	PLS	0.325	0.382
Hexa-	1	3-z Scales [30]	PLS	0.227	0.388
	2	5-z Scales [30]	PLS	0.223	0.434
	3	G-scales	PLS	0.267	0.555
Hepta-	1	3-z Scales [30]	PLS	0.164	0.589
	2	5-z Scales [30]	PLS	0.113	0.775
	3	G-scales	PLS	0.417	0.709
Octa-	1	3-z Scales [30]	PLS	0.391	0.586
	2	5-z Scales [30]	PLS	0.213	0.55
	3	G-scales	PLS	0.577	0.639
Deca-	1	3-z Scales [30]	PLS	0.238	0.719
	2	5-z Scales [30]	PLS	0.634	0.984
	3	G-scales	PLS	0.529	0.905

oligopeptides. Some insight into the relationships between structural features (such as the position of peptides and favorable amino acids) and biological activities of peptides was offered. G-scale proposed in this study is useful in structure characterization, and the results show that G-scale is superior to the previous. It has more advantages, such as definite physical and chemical meaning, and more abundant structure information and so on. G-scale may be a useful structural expression method for the study on QSAR of peptides, and it may hopefully be used in biological activity study of ACE-inhibitor oligopeptides.

Acknowledgments This study was supported by the National Natural Science Foundation of China (No. 60873103), and supported by the Key Project of Natural Science Foundation of China (No. 30830090), Create new drugs national major projects (2009ZX09503-005), and the Foundation for the Author of National Excellent Doctoral Dissertation of P.R. China (200776).

References

1. Phillips MI, Speakman EA, Kimura B (1993) Levels of angiotensin and molecular biology of the tissue rennin angiotensin systems. Regul Pept 43:1–20
2. Ondetti MA, Cushman DW (1982) Annu Rev Biochem 51:283–308
3. Hellberg S, Eriksson L, Jonsson J, Lindgren F, Sjostrom M, Skagerberg B, Wold S, Andrews P (1991) Int J Pept Prot Res 37:414–424
4. Bakulh HR, Shyam R, Asolekar (2001) QSAR models to predict effect of ionic strength on sorption of chlorinated benzenes and phenols at sediment-water interface. Water Res 35:3391–3401
5. Aleksandar Sablji C (2001) QSAR models for estimating properties of persistent organic pollutants required in evaluation of their environmental fate and risk. Chemosphere 43:363–375
6. Sneath PH (1966) Relations between chemical structure and biological activity in peptides. J Theor Biol 12:157–195
7. Kidera A, Konishi Y, Oka M et al (1985) A statistical analysis of the physical properties of the 20 naturally occuring amino acids. J Protein Chem 4:23–55
8. Hellberg S, Sjostrom M, Wold S (1986) The prediction of bradykinin potentiating potency of pentapep tides, an example of a peptide quantitative structure activity relationship. Acta Chem Scand B 40:135–140
9. Hellberg S, Sjostrom M, Skagerberg B et al (1987) Peptide quantitative structur activity relationships, a multivariate app roach. J Med Chem 30:1126–1135
10. Hellberg S, Eriksson L, Jonsson J et al (1991) Minimum analogue pep tidesets (MAPS) for quantitative structure-activity relationships. Int J Pept Protein Res 37:414–424
11. Sandberg M, Eriksson L, Jonsson J, Sjostrom M, Wold S (1998) New chemical descriptors relevant for the design of biologically active peptides, a multivariate characterization of 87 amino acids. J Med Chem 41:2481–2491
12. Collantes ER, Dunn WJ (1995) Amino acid side chain descrip tors for quantitative structure activity relationship studies of pep tide analogues. J Med Chem 38:2705–2713
13. Zaliani A, Gancia E (1999) MS-WHIM scores for amino acids: a new3D-descrip tion for pep tide QSAR and QSPR studies. J Chem Inf Comput Sci 39:525–533
14. Ramos De Armas R, Gonzalez-Diaz H, Molina R, Perez-Gonzalez M, Uriarte E (2004) Stochastic-based descriptors studying peptides biological properties: modeling the bitter tasting threshold of dipeptides. Bioorg Med Chem 12:4815–4822
15. Mei H, Liao ZH, Zhou Y, Li SSZ (2005) A new set of amino acid descriptors and its application in peptide QSARs. Biopolymers (Peptide Science) 80:775–786
16. Tian FF, Zhou P, Li ZL (2007) T-Scale as a novel vector of topological descriptors for amino acids and its application in QSARs of peptides. J Mol Struct Theochem 830:106–115
17. Tong J, Liu S, Zhou P, Wu B, Li Z (2008) A novel descriptor of amino acids and its application in peptide QSAR. J Theor Biol 253:90–97
18. Lin ZH, Long HX, Bo Z, Wang YQ, Wu YZ (2008) New descriptors of amino acids and their application to peptide QSAR study. Peptides 28:1798–1805
19. Shu M, Jiang Y, Yang L, Li Z et al (2009) Application of 'HESH' descriptors for the structure- activity relationships of antimicrobial peptides. Protein Pept Lett 16:143–149
20. Shu M, Mei H, Yang SB, Liao LM, Li ZL (2009) Structural parameter characterization and bioactivity simulation based on peptide sequence. QSAR Comb Sci 28:27–35
21. Yang L, Shu M, Jiang YJ, Mei H, Li Z (2010) ST-scale as a new set of amino acid descriptors and its application in QSAM of peptides and analogues. Amino Acids 38:805–816
22. Wold H (1985) Partial least squares in encyclopedis of statistical sciences. Wiley, New York, pp 581–591
23. Tenenhaus M (1998) La regression PLS theorie et pratique. Paris Editions Technip
24. Rose GD, Geslowitz AR, Lesser GJ, Lee RH (1985) Hydrophobicity of amino acid residues in globular proteins. Science 229:834–838
25. Wu JP, Aluko RE, Nakai S (2006) Structural requirements of angiotensin I-converting enzyme inhibitory peptides: quantitative structure–and–activity relationship study of di- and tripeptides. J Agric Food Chem 54:732–738
26. Cocchi M, Johansson E (1993) Amino acids characterization by GRID and multivariate data analysis. Quant Struct-Act Relat 12:1–8
27. Mei H, Zhou Y, Sun LL, Li ZL (2004) A new descriptors of amino acid and its application in peptide QSAR. Acta Phys Chim Sin 20:821–825
28. Zhou P, Zhou Y, Wu SR, Li B, Tian FF, Li ZL (2006) A new descriptor of amino acids based on the three-dimensional vector of atomic interaction field. Chin Sci Bull 51:524–529
29. Liang GZ, Zhou P, Zhou Y, Zhang QX, Li ZL (2006) New descriptors of amino acids and their applications to peptide quantitative structure activity relationship. Acta Chim Sinica 64:393–396
30. Wu JP, Aluko RE, Nakai S (2006) Structural requirements of angiotensin I-converting enzyme inhibitory peptides: quantitative structure-activity relationship modeling of peptides containing 4-10 amino acid residues. QSAR Comb Sci 10:873–880