Journal of the Serbian Chemical Society
JSCS-info@shd.org.rs • www.shd.org.rs/JSCS
J. Serb. Chem. Soc. 80 (3) 343–353 (2015)
JSCS–4720
UDC 547.466.1+539.16:547.466+532.74
Original scientific paper

A quantitative structure–activity relationship (QSAR) study of peptide drugs based on a new descriptor of amino acids

JIAN-BO TONG*, JIA CHANG, SHU-LING LIU and MIN BAI

Shaanxi University of Science and Technology, Xi'an 710021, China

(Received 4 June, accepted 7 July 2014)

Abstract: The quantitative structure–activity relationship (QSAR) approach is used for finding the relationships between molecular structures and the activity of peptide drugs. In this work, the stepwise multiple regression method was employed to select the optimal subset of descriptors that have significant contributions to the drug activity of 21 oxytocin analogues, 48 bitter tasting threshold activities and 58 angiotensin-converting enzyme inhibitors. A new set of descriptor, SVWGM, was used for the prediction of the activity of the peptide drugs and then were used to build a model by the partial least squares method, The stability and generalization ability of the model was strictly analyzed by both internal and external validations, by cross-validation and external validation correlations.

Keywords: amino acids; peptide; quantitative structure–activity relationships; partial least square; SVWGM.

INTRODUCTION

As one of the hottest topic, peptide drugs, which perform important roles in all living systems, have been widely debated.[1,2] Most of experimental methods on peptide drugs were inefficient and expensive and hence, computational methods, such as quantitative structure–activity relationships (QSAR), which involve not only the key idea of pharmacology, but also the foundation of drug design, were brought into the spotlight. In a QSAR study of peptides, the major information on structure and function for peptides is contained in their amino acid sequence and hence it is crucial to characterize the sequence structure. The pioneering work on amino acid descriptors was undertaken by Sneath *et al.*[3] who used semi-quantitative physicochemical data to derive descriptors for the 20 naturally occurring amino acids. Kidera *et al.*[4–6] collected 188 properties of the 20 natural amino acids and employed factor analysis to obtain 10 orthogonal factors that were the most important for determining the three-dimensional structure

* Corresponding author. E-mail: jianbotong@aliyun.com
doi: 10.2298/JSC140604069T

of proteins. Soon afterwards, Hellberg *et al.*[5,7–12] extracted three principal components (PCs), or *z*-scales, from 29 physicochemical properties of coded amino acids by PC analysis (PCA). In recent years, a new series of amino acid indices based on quantum topological molecular similarity (QTMS) descriptors[12,13] were proposed for peptide QSAR studies and good results were obtained. The study of the structure-based approaches for peptides by Hou *et al.*[14–17] achieved satisfactory results. Since then, many quantitative descriptors for amino acid were proposed and successfully used in practice.[18,19] They were then applied to peptide QSAR analyses and showed good structural characterization ability.

PRINCIPLE AND METHODS

Molecular structural characterization

Molecular structural characterization (MSC) focuses on transforming molecular structural characteristics into a group of numerical codes, dedicated to minimizing information loss during this process. In this study, 99 weighted holistic invariant molecular (WHIM) descriptors (obtained as statistical indices of the atoms projected onto the 3 principal components obtained from weighted covariance matrices of the atomic coordinates, the meanings of which are listed in Table S-I of the Supplementary material to this paper), 74 geometrical descriptors (obtained from molecular geometry, the meanings are listed in Supplementary Table S-II of the Supplementary material) and 160 3D-molecule representations of structures based on electron diffraction (3D-MORSE) descriptors (calculated by summing atom weights viewed by a different angular scattering function, the meanings of which are listed in Supplementary Table S-III of the Supplementary material) for each single amino acid were generated *via* Dragon 5.4 software.

The original variable matrices $X_{20\times99}$ (Table S-IV of the Supplementary material), $X_{20\times74}$ (Table S-V of the Supplementary material) and $X_{20\times160}$ (Table S-VI of the Supplementary material were submitted to PCA, autoscaling was always applied. The PCA scores (Table S--VII of the Supplementary material) were extracted for all those components that were identified to be necessary to explain a minimum of 85 % of the total variance, also, the corresponding eigenvalue > 1. A total of 18 PCs (Table S-VIII of the Supplementary material) were selected from the PCA for each amino acid, which were treated as a new vector descriptor – SVWGM (principal component scores vector of WHIM descriptors, Geometrical descriptors and 3D-MORSE descriptors) to characterize a peptide. PCA was implemented by MATLAB R2010a software.

Variable selection and modeling

The selection of molecular descriptors most relevant for acute drug activity is a key problem involved in QSAR, as is the choice of a proper technique for the construction of models. At present, the stepwise multiple regression (SMR) method is an interesting and widely used variable selection method.[20] SMR was implemented by SPSS software for Windows 16.0.

The partial least squares (PLS) method was employed for developing a reliable and predictive QSAR model and for an estimation of its stability, which was strictly analyzed by the cross-validation correlation coefficient (Q^2_{LOO}) and correlation coefficient (R^2_{cum}). The PLS analyses were implemented using SIMCA-P 12.0 software.

RESULTS AND DISCUSSION

QSAR model for oxytocin

Oxytocin (OT), a neurohypophyseal hormone, stimulates contraction of the uterine myometrium at parturition and contracts myoepithelial cells during lactation. Twenty-one oxytocin analogues[21] varying at positions 2, 3 and 8 are often utilized to test the effectiveness of diverse kinds of amino acid descriptors for its oxytocin-activity denoted by *OA* (Table I). For each peptide, its structure was first quantified by 3×18 SVWGM scales. In order to prove the validity and stability of the model, the whole data set was systematically divided into two subsets. The dataset (n = 21) was divided into a training set of 18 samples and a test set (highlighted with "*" in Table I) of 3 samples. The training sets were then used for developing QSAR models employing the PLS technique, while the corresponding test set was utilized to evaluate the predictive ability of the obtained model.[12] In the PLS regression analysis, the Q^2_{LOO} values (Table S-IX of the Supplementary material) changed with the SMR-introduced variables, achieving the maximum (0.864) at the fifth step, meanwhile, the value of the correlation coefficient (R^2_{cum}) reached up to 0.944. The optimum PLS model with 5 variables (v_{24}, v_{17}, v_{36}, v_{51} and v_{23}) was then obtained, wherein only two PLS components were sufficient to account for 97.2 % of the variances of the Y variables. A plot of the observed values for OT *vs*. the predicted ones (listed in Table I) is presented in Fig. 1, from which it could be seen that the points were minutely scattered about the through the origin-passing diagonal and the model possessed a strong fitting capability. The results obtained from the PLS models using the SVWGM descriptors are given in Table II, which were superior to those from the literature (Table II). A scoring distribution scatter of the 21 samples in the first and second PLS principal component spaces is shown in Fig. 2, from which it could be seen that most samples fell within the Hotelling T^2 confidence ellipse[22,23] with a 95 % confidence interval, with the exception of $7^{\#}$.

The distance to the PLS model in the X space is described in Fig. 3. It can be seen that the normalized distance to X for none of sample was beyond the critical value of 2.148 (significance level = 5 %).

TABLE I. Amino acid sequences of 21 OT and their observed and predicted activities; Obs – observed values, Pred – predicted values, * – test set

No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred
1	YYL	2.00	2.54	8	SIK	1.00	1.03	15	YFL	4.30	4.22
2	FFL	3.52	3.60	9	YWK	1.00	0.82	16	YII	5.46	5.12
3	YWL	1.60	1.90	10	FIK	3.00	3.52	17	YFK	3.70	3.13
4	YLL	4.65	5.27	11	YIR	4.88	5.11	18	YYK*	1.00	1.45
5	YVL	4.77	4.95	12	YFH*	3.18	3.09	19	FFK	2.48	2.51
6	YIV*	5.30	4.26	13	YIL	5.65	5.24	20	YIK	4.89	4.15
7	FYK	1.00	0.83	14	FIL	4.50	4.62	21	YFR	4.30	4.10

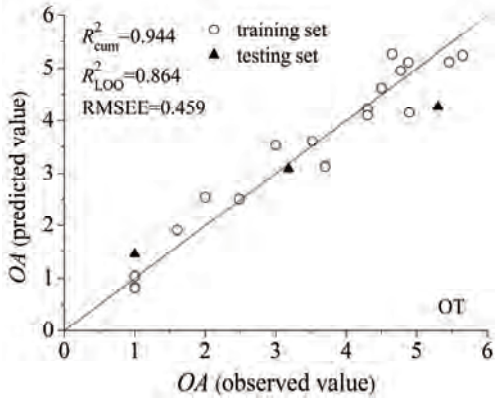

Fig. 1. Plot of predicted *vs.* observed values for 21 OT by the PLS model.

TABLE II. Comparison between the different QSAR models of OT; –: not determined

No.	Descriptors	Model	A[a]	R^2_{cum} [b]	Q^2_{LOO} [c]	*RMSEE*[d]
1	SVRDF	PLS	2	0.941	0.811	0.417
2	DISMAT 250	SMR	3	0.931	0.907	–
3	**SVWGM**	**PLS**	**2**	**0.944**	**0.864**	**0.459**
4	SVRG	PLS	2	0.955	0.908	0.368
5	SVRG	PLS	3	0.968	0.931	0.317

[a]principal components, [b]cumulative multiple correlation coefficient, [c]cumulative cross-validated, [d]root mean square error

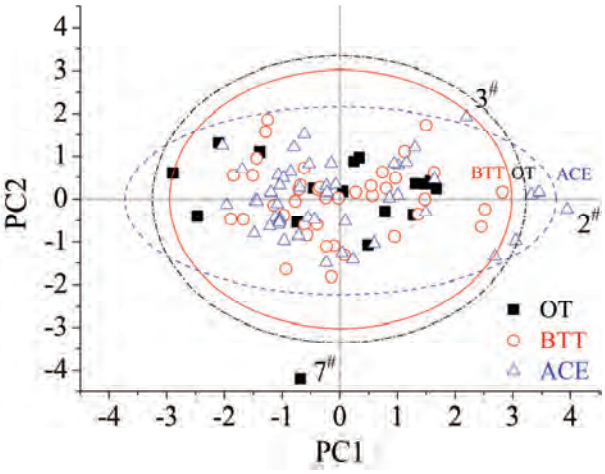

Fig. 2. Score scatter of the top two principle components for 21 OT, 48 BTT and 58 ACE.

A good statistical result often indicates a good robustness and high internal predictive power of a QSAR model. The cross-validation coefficient between the predicted and the observed values of the test set, Q^2_{ext}, was used to verify the external predictive ability of the PLS model. Recently, several novel methods for model validation were developed, such as the leave-several-out (LSO),[24]

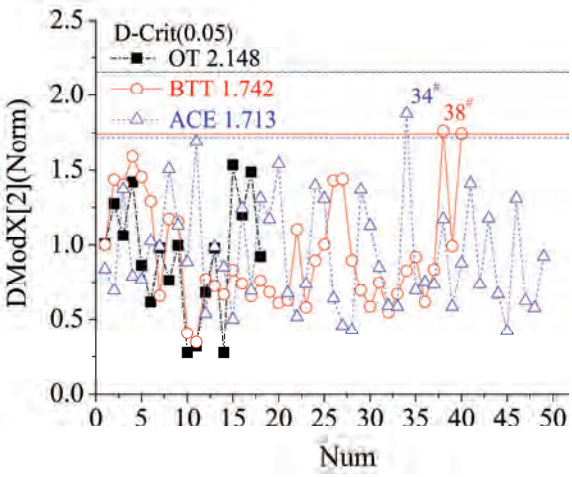

Fig. 3. Distance to the PLS model in *X* space of 21 OT, 48 BTT and 58 ACE.

Y-scrambling,[25] self-organizing mapping of molecular objects,[26] and external validation using division of a dataset into training and test sets.[25–27] In this work, the external predictive ability of the PLS model was calculated according to Eq. (1):[12,25]

$$Q^2_{\text{ext}} = 1 - \frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{\sum (Y_{\text{obs}} - \overline{Y}_{\text{training}})^2} \quad (1)$$

where Y_{obs} and Y_{pred} indicate, respectively, the values of observed and predicted activities (the predicted values are listed in Table I), while $\overline{Y}_{\text{training}}$ indicates the mean activity value of the training set. The constructed model was then utilized to predict the test set, which gave Q^2_{ext} value of 0.863. All the results showed that the presented PLS model had an excellent prediction capability for drug design.

QSAR model for bitter tasting thresholds

As a classical sample set in QSAR studies, the 48 bitter tasting thresholds (BTT) reported by Collantes,[28] with their activities denoted by –log IC_{50} values (Table III), are often utilized to validate the efficiency of amino acid descriptors. Accordingly, each BTT could be characterized with 2×18 SVWGM scales, and using the same method, the 48 BTT were divided into 40 training samples and 8 test samples (highlighted with "*" in Table III). The optimum QSAR models of –log (IC_{50}) were created by the SMR–PLS method with 7 independent variables (v_{32}, v_1, v_{34}, v_{36}, v_{25}, v_6 and v_{12}). The results showed that the established model had a good predictive ability and strong robustness, with an R^2_{cum} value of 0.911, a Q^2_{LOO} of 0.818 (Table S-X of the Supplementary material), and Q^2_{ext} of 0.825. The plot of predicted value (predicted values are listed in Table III) against the observed data for BTT is presented in Fig. 4, from which it could be seen that

most samples were shown to be uniformly dispersed around the through the origin-passing diagonal. The results obtained from the PLS models using SVWGM descriptors are given in Table IV, showing that the herein obtained results were superior to those reported in the literature (Table IV). The scoring distribution scatter of the 48 samples in the spaces of the first and second PLS principal component is presented in Fig. 2. It could be seen that all samples fell within the Hotelling T^2 95 % confidence ellipse. The distance to the PLS model in the *X* space is shown in Fig. 3, from which it could be seen that the normalized distance to *X* for most sample was smaller than the critical value of 1.742 (significance level = 5 %), except for sample 38#.

TABLE III. Amino acid sequences of 48 BTT and their observed and predicted activities; Obs – observed values, Pred – predicted values, * – test set

No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred
1	GV	1.13	0.96	17	LL	2.35	2.55	33	IS	1.49	1.59
2	GL	1.68	1.54	18	LF*	2.75	2.88	34	IT	1.49	1.87
3	GI	1.70	1.60	19	LW	3.40	3.12	35	PA	1.32	1.39
4	GP	1.35	1.42	20	LY	2.46	2.67	36	PL*	2.22	2.09
5	GF	1.80	1.87	21	IG	1.68	1.40	37	PI	2.33	2.15
6	GW*	1.89	2.10	22	IA	1.68	1.73	38	PY	1.80	2.21
7	GY	1.77	1.66	23	IV	2.05	1.85	39	PF	2.80	2.42
8	AV	1.16	1.07	24	IL*	2.26	2.42	40	FG	1.77	1.75
9	AL	1.70	1.64	25	II	2.26	2.48	41	FL	2.87	2.78
10	AF	1.72	1.97	26	IP	2.40	2.31	42	FP*	2.70	2.67
11	VG	1.19	1.24	27	IW	3.05	2.99	43	FF	3.10	3.11
12	VA*	1.16	1.57	28	IN	1.49	1.69	44	FY	3.13	2.91
13	VV	1.71	1.69	29	ID	1.37	1.46	45	WE	1.56	1.50
14	VL	2.00	2.26	30	IQ*	1.49	1.17	46	WW	3.60	3.36
15	LG	1.72	1.53	31	IE	1.37	1.13	47	YL	2.40	2.54
16	LA	1.72	1.86	32	IK	1.65	1.84	48	SL*	1.49	1.68

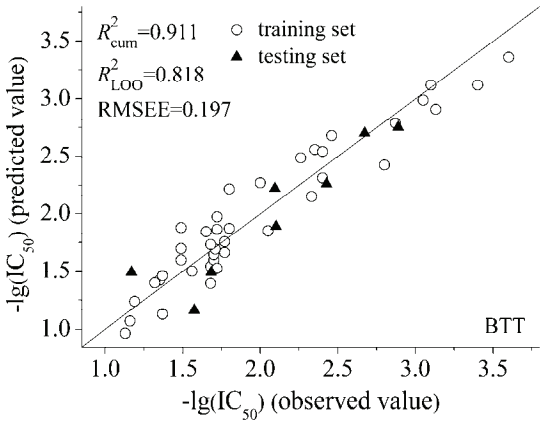

Fig. 4. Plot of predicted *vs*. observed values for 48 BTT by the PLS model.

TABLE IV. Comparison between QSAR models of BTT; –: not determined

No.	Descriptor	Model	A[a]	R^2_{cum} [b]	R^2_{LOO} [c]	*RMSEE*[d]
1	MEEV	MLR	3	0.649	0.57	0.37
2	MS–WHIM	PLS	3	0.704	0.633	–
3	MEEV	MLR	10	0.711	0.475	0.34
4	MEEV	MLR	3	0.735	0.677	0.32
5	MS–WHIM	PLS	3	0.754	0.71	0.32
6	SVRDF	PLS	1	0.766	0.86	0.209
7	MEEV	MLR	10	0.773	0.588	0.33
8	z scale	PLS	2	0.824	–	0.26
9	SSIA–HF	PLS	2	0.844	0.798	0.25
10	T-scales	PLS	2	0.845	0.786	0.39
11	c-scales	PLS	3	0.847	0.863	0.195
12	ISA–ECI	PLS	2	0.847	–	–
13	SSIA–AM1	PLS	1	0.85	0.837	0.25
14	SSIA–DFT	PLS	2	0.856	0.741	0.24
15	VSW	PLS	2	0.868	0.696	0.24
16	VSW	PLSd	2	0.873	0.751	0.23
17	VHSE	PLS	3	0.881	0.843	0.22
18	SVGETAWAY	PLS	2	0.887	0.753	0.216
19	SSIA–PM3	PLS	4	0.888	0.829	0.22
20	SZOTT	PLS	2	0.908	0.736	0.2
21	SVRG	PLS	3	0.91	0.863	0.195
22	**SVWGM**	**PLS**	**2**	**0.911**	**0.818**	**0.197**
23	SVHEHS	MLR	6	0.949	0.886	0.136
24	GRID	PLS	1	–	0.78	–

[a]principal components, [b]cumulative multiple correlation coefficient, [c]cumulative cross-validated, [d]root mean square error

QSAR model for angiotensin-converting enzyme inhibitors

The angiotensin-converting enzyme (ACE) plays a critical physiological role in raising blood pressure, the inhibition of ACE activity could lead to an overall antihypertensive effect.[29–31] Dipeptide sequences of 58 ACE inhibitors, a classical sample set for QSAR studies, with their activities expressed as log ($1/IC_{50}$) values,[10,32] was employed. The structure of each peptide was first quantified by 2×18 SVWMG scales. As shown in Table V, the dataset (n = 58) was divided into a training set of 49 samples and test set (highlighted with "*" in Table V) of 9 samples. The training set was then used for the development of QSAR models using the PLS technique, while the corresponding test set was utilized to validate the external prediction power of the developed model. In PLS regression analysis, Q^2_{LOO} (Table S-XI of the Supplementary material) changed with the SMR--introduced variables, achieving a maximum (0.799) at the seventh step, meanwhile, the correlative coefficient (Q^2_{cum}) was up to 0.865. The optimum PLS model with 7 variables (v_{23}, v_{34}, v_6, v_{28}, v_{30}, v_9 and v_{21}) was then obtained, wherein only two significant principal components were necessary to explain

TABLE V. Amino acid sequences of 58 ACE and their observed and predicted activities; Obs – observed values, Pred – predicted values, * – test set

No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred	No.	Peptide	Obs	Pred
1	VW	5.80	5.19	21	IG	2.92	2.93	41	GG	2.14	2.13
2	IW	5.70	5.58	22	GI	2.92	2.72	42	QG*	2.13	2.27
3	IY	5.43	4.48	23	GM	2.85	2.62	43	SG	2.07	2.30
4	AW	5.00	5.09	24	GA*	2.70	2.54	44	LG	2.06	2.57
5	RW	4.80	5.28	25	YG	2.70	2.39	45	GD	2.04	2.41
6	VY*	4.66	4.08	26	GL	2.60	3.10	46	TG	2.00	2.31
7	GW	4.52	4.79	27	AG	2.60	2.36	47	EG	2.00	2.00
8	VF	4.28	3.72	28	GH	2.51	2.40	48	DG*	1.85	1.99
9	AY	4.06	3.91	29	GR	2.49	2.41	49	PG	1.77	1.78
10	IP	3.89	4.02	30	KG*	2.49	2.27	50	LA	3.51	2.98
11	RP	3.74	3.71	31	FG	2.43	2.45	51	KA	3.42	2.68
12	AF*	3.72	3.55	32	GS	2.42	2.92	52	RA	3.34	3.03
13	GY	3.68	3.67	33	GV	2.34	2.46	53	YA	3.34	2.80
14	AP	3.64	3.46	34	MG	2.32	2.93	54	AA*	3.21	2.78
15	RF	3.64	3.81	35	GK	2.27	1.83	55	FR	3.04	2.74
16	VP	3.38	3.63	36	GE*	2.27	2.00	56	HL	2.49	3.13
17	GP	3.35	3.22	37	GT	2.24	2.39	57	DA	2.42	2.41
18	GF*	3.20	3.32	38	WG	2.23	2.14	58	EA	2.00	2.41
19	IF	3.03	4.12	39	HG	2.20	2.17				
20	VG	2.96	2.53	40	GQ	2.15	2.63				

93.0 % of the variance of *Y*. The model exhibited good prediction ability and stability, which was better than the models established by other methods (Table VI). The plot predicted by the optimum QSAR models and observed values are shown in Fig. 5 (the predicted values are listed in Table VI), where most samples were shown to be uniformly dispersed around the through the origin-passing diagonal and the model possessed a strong capability for fitting. The scoring distribution scatter of the 58 samples in the first and second PLS principal component spaces is presented in Fig. 2. It could be seen that all samples fell within the Hotelling T^2 95 % confidence level ellipse, with the exception of $2^{\#}$ and $3^{\#}$. The errors may be caused by the characterization method, experiment, or other reasons. No matter what kinds of reasons, there are only two samples with large errors, and the excellent results of this study should be recognized. The distance to the PLS model in the *X* space was described in Fig. 3. It could be seen that the normalized distance to *X* for most samples was smaller than the critical value of 1.713 (significance level = 5 %), except for sample $34^{\#}$.

In the optimum model, the external prediction power was evaluated using Eq. (1) and the Q^2_{ext} value reached 0.883. All the results showed that the presented PLS model could be effectively used to predict the acute activity of ACE.

TABLE VI. Comparison between QSAR models of ACE; –: not determined

No.	Descriptor	Model	A[a]	R^2_{cum} [b]	R^2_{LOO} [c]	*RMSEE*[d]
1	MEEV	MLR	10	0.711	0.475	0.34
2	MEEV	MLR	3	0.649	0.57	0.37
3	MS–WHIM	PLS	3	0.657	0.541	–
4	MS–WHIM	PLS	3	0.708	0.637	0.54
5	SSIA–DFT	PLS	2	0.734	0.678	0.52
6	MEEV	MLR	3	0.735	0.677	0.32
7	SSIA–AM1	PLS	1	0.769	0.699	0.49
8	z-scale	PLS	2	0.77	0.723	–
9	VHSE	PLS	1	0.77	0.745	0.48
10	MEEV	MLR	10	0.773	0.588	0.33
11	SVTV	PLS	1	0.789	0.767	0.45
12	VSTV	PLS	1	0.789	0.767	0.46
13	SSIA–PM3	PLS	4	0.789	0.773	0.47
14	SSIA–HF	PLS	2	0.797	0.76	0.46
15	T-scale	PLS	2	0.845	0.786	0.39
16	HSEHPCSV	PLS	1	0.846	0.835	0.396
17	V	PLS	2	0.849	0.783	0.4
18	VSW	PLSd	1	0.861	0.835	0.38
19	**SVWGM**	**PLS**	**2**	**0.865**	**0.799**	**0.386**
20	VSW	PLS	2	0.868	0.784	0.37
21	MHDV	PCR	19	0.878	0.753	0.35
22	SVWG	PLS	3	0.893	0.83	0.34
23	SZOTT	PLS	2	0.894	0.828	0.331
24	SVEEVA	PLS	2	0.894	0.839	0.332
25	SVHEHS	MLR	8	0.936	0.854	0.259

[a]principal components, [b]cumulative multiple correlation coefficient, [c]cumulative cross-validated, [d]root mean square error

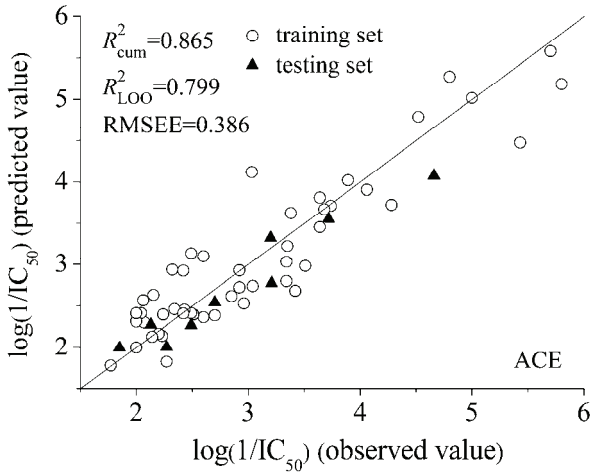

Fig. 5. Plot of predicted *vs.* observed values for 58 ACE by the PLS model.

CONCLUSIONS

In this study, a new set of descriptors-SVWGM was derived from principal component analyses of 99 WHIM descriptors, 74 geometrical descriptors and 160 3D-MORSE descriptors of 20 natural amino acids. On applying the SVWGM scales in the peptide QSAR studies for three kinds of classical peptide analogues, the results were similar to, or better than those found in the literature. The satisfactory results showed that the models had good prediction capability and favorable stability and the SVWGM descriptors could be well used to characterize molecular structure information and express the QSAR of peptide drugs.

SUPPLEMENTARY MATERIAL

The meanings and values of descriptors are available electronically from http:// //www.shd.org.rs/JSCS/, or from the corresponding author on request.

Acknowledgements. This work was supported by the National Natural Science Funds of China (21475081) and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

И З В О Д

QSAR СТУДИЈА ПЕПТИДНИХ ЈЕДИЊЕЊА ЗАСНОВАНА НА НОВОМ ДЕСКРИПТОРУ АМИНОКИСЕЛИНА

JIAN-BO TONG, JIA CHANG, SHU-LING LIU и MIN BAI

Shaanxi University of Science & Technology, Xi'an, China

Метода квантитативних релација између структуре и активнисти (QSAR) примењена је за налажење везе између молекулске структуре и активности пептидних једињења. У овом раду коришћена је вишестепена вишеструка регресија да би се изабрао оптимални подскуп дескриптора, који имају значајан допринос активности за 21 окситоцински аналог, 48 једињења горког укуса и 58 инхибитора ангиотензин-конвертујућег ензима. Примењен је нови сет дескриптора SVWGM. Помоћу њих је добивен модел за предвиђање активности испитиваних супстанци. Модел је конструисан применом методе парцијалних најмањих квадрата, а затим је анализиран одговарајућим статистичким тестовима.

(Примљено 4. јуна, прихваћено 7. јула 2014)

REFERENCES

1. J. Jirácek, A. Yiotakis, B. Vincent, A. Lecoq, A. Nicolaou, F. Checler, *J. Biol. Chem.* **270** (1995) 21701
2. M. Marraud, A. Aubry, *J. Pept. Sci.* **40** (1996) 45
3. P. H. A. Sneath, *J. Theor. Biol.* **12** (1966) 157
4. A. Kidera, Y. Konishi, M. Oka, T. Ooi, H. A. Scheraga, *J. Protein Chem.* **4** (1985) 23
5. A. Kidera, Y. Konishi, T. Ooi, H. A. Scheraga, *J. Protein Chem.* **4** (1985) 265
6. K. Nakai, A. Kidera, M. Kanehisa, *Protein Eng.* **2** (1988) 93
7. S. Hellberg, L. Eriksson, J. Jonsson, F. Lindgren, M. Sjöström, B. Skagerberg, S. Wold, P. Andrews, *Int. J. Pept. Protein Res.* **37** (1991) 414
8. S. Hellberg, M. Sjöström, B. Skagerberg, S. Wold, *J. Med. Chem.* **30** (1987) 1126
9. S. Hellberg, M. Sjöström, S. Wold, *Acta Chem. Scand., B* **40** (1986) 135

10. S. Wold, L. Eriksson, S. Hellberg, J. Jonsson, M. Sjöström, B. Skagerberg, C. Wikström, *Can. J. Chem.* **65** (1987) 1814
11. J. Jonsson, L. Eriksson, S. Hellberg, M. Sjöström, S. Wold, *Quant. Struct.-Act. Relat.* **8** (1989) 204
12. L. Eriksson, J. Jonsson, S. Hellberg, F. Lindgren, B. Skagerberg, M. Sjostrom, S. Wold, *Acta Chem. Scand.* **44** (1990) 50
13. B. Hemmateenejad, S. Yousefinejad, A. R. Mehdipour, *Amino Acids* **40** (2011) 1169
14. T. J. Hou, Z. Xu, W. Zhang, W. A. McLaughlin, D. A. Case, Y. Xu, W. Wang, *Mol. Cell. Proteomics* **8** (2009) 639
15. T. J. Hou, Y. Y. Li, W. Wang, *Bioinformatics* **27** (2011) 1814
16. T. J. Hou, N. Li, Y. Y. Li, W. Wang, *J. Proteome Res.* **11** (2012) 2982
17. Z. Xu, T. J. Hou, N. Li, Y. Xu, W. Wang, *Mol. Cell. Proteomics* **11** (2012) O111
18. M. Sandberg, L. Eriksson, J. Jonsson, M. Sjöström, S. Wold, *J. Med. Chem.* **41** (1998) 2481
19. X. Y. Wang, J. Wang, Y. Lin, Y. Ding, Y. Q. Wang, X. M. Cheng, Z. H. Lin, *J. Mol. Model.* **17** (2011) 1599
20. A. Jain, D. Zongker, *IEEE Trans. Pattern Anal. Mach. Intell.* **19** (1997) 153
21. H. Mei, Y. Zhou, L. L. Sun, Z. L. Li, *Acta Phys.-Chim. Sin.* **20** (2004) 821
22. J. P. George, Z. Chen, P. Shaw, *World Acad. Sci. Eng. Technol.* **50** (2009) 970
23. T. Iwashita, *J. Statist. Plan. Inf.* **61** (1997) 85
24. J. Polanski, R. Gieleciak, A. Bak, *Comb. Chem. High. T. Scr.* **7** (2004) 793
25. A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **22** (2003) 69
26. J. Polanski, A. Bak, R. Gieleciak, T. Magdziarz, *J. Chem. Inf. Model.* **46** (2006) 2310
27. A. Golbraikh, A. Tropsha, *J. Mol. Graphics Modell.* **20** (2002) 269
28. E. R. Collantes, W. J. Dunn, *J. Med. Chem.* **38** (1995) 2705
29. R. Natesh, S. L. Schwager, E. D. Sturrock, K. R. Acharya, *Nature* **421** (2003) 551
30. Z. P. Yu, B. Q. Liu, W. Z. Zhao, Y. G. Yin, J. B. Liu, F. Chen, *Food Chem.* **133** (2012) 315
31. M. A. Ondetti, D. W. Cushman, *Annu. Rev. Biochem.* **51** (1982) 283
32. F. F. Tian, P. Zhou, Z. L. Li, *J. Mol. Struct.* **830** (2007) 106.