

UDC 539.2

QSAR STUDY OF PEPTIDE DRUGS BY 3D-HoVAIF

J. Tong, J. Chang, L. Li, M. Bai

College of Chemistry and Chemical Engineering, Shaanxi University of Science & Technology, Xi'an, P.R. China
E-mail: jianbotong@aliyun.com (J.B. Tong)

Received April, 26, 2014

By applying a three-dimensional holographic vector of the atomic interaction field (3D-HoVAIF) to express the structure of three classical peptide drugs, quantitative structure activity relationship (QSAR) models are built by the multiple linear regression. The accuracy of the proposed model is illustrated using Q_{LOO}^2 (cross-validation) and r^2 (test set validation). Moreover, the r_m^2 metrics is used to further refine the predictive ability of the developed QSAR models. The results show that 3D-HoVAIF, due to the high predictive ability, offers a useful alternative to the costly and time-consuming experiments determining the bioactivity of peptide drugs.

DOI: 10.15372/JSC20150706

Keywords: peptide drugs, 3D-HoVAIF, QSAR, MLR, r_m^2 .

The hottest topic of peptide drugs which play an important role [1, 2] in all living systems have been widely debated. Most experimental methods on peptide drugs were inefficient and expensive, while the computational methods such as QSAR have been brought into the spotlight. They consider the key idea of pharmacology and underlie the drug design. In the QSAR study of peptides, major information about the structure and function of 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 physicochemical semi-quantative data to derive descriptors for the 20 naturally occurring amino acids. Kidera *et al.* [4, 5] collected 188 properties of the 20 natural amino acids and employed the factor analysis to obtain 10 orthogonal factors that are most important for determining the three dimensional structure of proteins. Soon afterwards, Hellberg *et al.* [5–9] extracted three principal components (PCs), or z-scales, from 29 physicochemical properties of coded amino acids by the principal component analysis (PCA). In the recent year, a new series of amino acid indices based on quantum topological molecular similarity (QTMS) descriptors [10, 11] have been proposed for peptide QSAR studies and yielded a good result. Since then, many amino acid quantitative descriptors have been successfully proposed and used in practice [12, 13]. They were then applied to the peptide QSAR analysis and showed good structural characterization ability.

Based on the previous work, a novel 3D molecular structural characterization method, the three-dimensional holographic vector of the atomic interaction field (3D-HoVAIF) [14, 15], which included the merits of traditional 2D and 3D descriptors for it does not require conformation alignment, was derived from multi-dimensional vectors to represent molecular steric structural characteristics. In this study, 3D-HoVAIF was used to express the structure of amino acid molecules and for further research of peptide QSAR.

MATERIALS AND METHODS

Peptide data sets. For the present work we have used three different datasets for the QSAR modeling of peptides: (i) bitter tasting thresholds (BTT) of 48 dipeptides were selected by Collantes [16], their activities are expressed as a negative logarithm of bitter-tasting threshold concentrations (pT); (ii) oxytocin (OT) of 21 tripeptides [3, 17] varying at positions 2, 3, and 8; they are often utilized to the test effectiveness of diverse kinds of amino acid descriptors for the oxytocin activity denoted by OA [3]; and (iii) a dipeptide sequence of 58 angiotensin-converting enzyme (ACE) inhibitors [6–9, 16, 18]. Each peptide biological activity was expressed in the form of pIC_{50} . The datasets are shown in Table 1.

Three-dimensional holographic vector of the atomic interaction field. 3D-HoVAIF was developed considering three common atomic non-covalent bond interactions of biological activities, i.e., electrostatic, steric, and hydrophobic interactions related to the atomic relative distance and atomic self-properties. These descriptors neither resort to any experimental parameters nor consider the configurational overlap of samples. Ordinary atoms of organic molecules, including H, C, N, P, O, S, F, Cl, Br, and I, are classified into five types in the Periodic Table of Elements. According to the hybridization state of atoms, the atoms are furthermore subdivided into ten types. Thus, there are 55 interatomic interactions (Supplementary Table 1) in a molecule. For specific details see the previous articles [14, 15].

In this study, electrostatic, steric, and hydrophobic potential energies take part in the representation of different interactions, producing $3 \times 55 = 165$ interaction items for amino acid molecules. Since there are no halogen atoms as the fifth atoms and sp hybridized C, N, etc. in amino acid molecules, 84 all-zero-terms having no statistical significance appear and we get rid of these variables. Each amino acid molecule is characterized by the other 81 3D-HoVAIFs (shown in bold in Supplementary Table 1).

Feature selection. In the QSAR study, the feature selection refers to a procedure which selects a subset of descriptors from a pool of candidate descriptors. This procedure increases the reliability and reduces the cost and time of the modeling process [19]. In this study, the stepwise multiple regression (SMR) was carried out to select the variables.

MLR modeling and model validation. The multiple linear regression [20, 21] (MLR) is a classical mathematical modeling method to obtain a more detailed insight into the structure-activity relationships between the molecular structure and bioactivity. The MLR analysis has been carried out to derive the best QSAR model.

A rigorous validation of the developed models plays the key role in their successful application for the prediction of new peptide drugs. In this study, the model was selected based on the leave-one-out (LOO) cross-validation correlation coefficient (Q_{LOO}^2) for the internal validation and the correlation coefficient between the predicted and observed activities (r^2) for the external validation. Moreover, a novel metrics r_m^2 is used to further refine the predictive ability of the developed QSAR models. It was introduced by Roy and coworkers [22, 23] and determined the proximity between the observed and predicted response data, denoting the ability of the model to ideally predict the test set molecules. Models are considered acceptable if they satisfy all of the following conditions [24, 25]:

$$Q_{LOO}^2 = 1 - \sum (Y_{obs} - Y_{pred})^2 / \sum (Y_{obs} - \bar{Y}_{training})^2 \quad (Q_{LOO}^2 > 0.6), \quad (1)$$

$$r^2 = [\sum (Y_{obs} - \bar{Y}_{obs})(Y_{pred} - \bar{Y}_{pred})]^2 / \sum (Y_{pred} - \bar{Y}_{pred})^2 \sum (Y_{obs} - \bar{Y}_{obs})^2 = \quad (r^2 > 0.6), \quad (2)$$

$$= [\sum (Y_{pred} - \bar{Y}_{pred})(Y_{obs} - \bar{Y}_{obs})]^2 / \sum (Y_{obs} - \bar{Y}_{obs})^2 \sum (Y_{pred} - \bar{Y}_{pred})^2$$

$$r_0^2 = 1 - \sum (Y_{obs} - kY_{pred})^2 / \sum (Y_{obs} - \bar{Y}_{obs})^2 \quad ((r^2 - r_0^2)/r^2 < 0.1), \quad (3)$$

$$r_0'^2 = 1 - \sum (Y_{pred} - k'Y_{obs})^2 / \sum (Y_{pred} - \bar{Y}_{pred})^2 \quad ((r^2 - r_0'^2)/r^2 < 0.1), \quad (4)$$

$$r_m^2 = r^2 (1 - \sqrt{r^2 - r_0^2}) \quad (r_m^2 > 0.5), \quad (5)$$

$$r_m'^2 = r^2 (1 - \sqrt{r^2 - r_0'^2}) \quad (r_m'^2 > 0.5), \quad (6)$$

Table 1

Amino acids sequences of three different peptide drugs and their activities

	Sample	Obsd	Pred ^a	Pred ^b	Sample	Obsd	Pred ^a	Pred ^b	Sample	Obsd	Pred ^a	Pred ^b	
BTT	GV ^{a,b}	1.13	1.29	1.12	LL ^{a,b}	2.35	2.53	2.38	IS ^a	1.49	1.31	1.35	
	GL	1.68	1.62	1.48	LF	2.75	2.87	2.77	IT	1.49	1.46	1.42	
	GI	1.7	1.59	1.43	LW	3.4	3.11	3.12	PA	1.32	1.32	1.29	
	GP	1.35	1.45	1.47	LY	2.46	2.54	2.54	PL	2.22	2.16	2.08	
	GF ^a	1.8	1.97	1.86	IG ^{a,b}	1.68	1.58	1.69	PI ^{a,b}	2.33	2.13	2.03	
	GW	1.89	2.19	2.00	IA	1.68	1.65	1.66	PY	1.8	2.17	2.12	
	GY ^b	1.77	1.63	1.63	IV	2.05	2.15	2.08	PF	2.8	2.5	2.46	
	AV	1.16	1.1	1.09	IL	2.26	2.48	2.44	FG	1.77	1.95	1.98	
	AL ^a	1.7	1.43	1.45	II ^a	2.26	2.45	2.4	FL ^{a,b}	2.87	2.85	2.73	
	AF	1.72	1.77	1.83	IP	2.4	2.32	2.43	FP	2.7	2.68	2.72	
	VG ^b	1.19	1.13	1.37	IW ^b	3.05	3.06	3.17	FF	3.1	3.19	3.11	
	VA	1.16	1.19	1.34	IN	1.49	1.42	1.37	FY	3.13	2.86	2.88	
	VV ^a	1.71	1.69	1.76	ID ^a	1.37	1.3	1.34	WE ^a	1.56	1.88	1.69	
	VL	2	2.02	2.12	IQ	1.49	1.57	1.46	WW	3.6	3.56	3.49	
	LG	1.72	1.63	1.64	IE ^b	1.37	1.38	1.37	YL ^b	2.4	2.46	2.7	
	LA	1.72	1.7	1.61	IK	1.65	1.65	1.83	SL	1.49	1.38	1.65	
	OT	YYL ^{a,b}	2	2.15	1.91	SIK	1	1.02	0.91	YFL	4.3	4.35	4.39
		FFL	3.52	3.39	3.21	YWK ^a	1	0.93	0.89	YII	5.46	5.27	5.49
YWL		1.6	1.64	1.7	FIK	3	3.63	3.54	YFK ^{a,b}	3.7	3.64	3.57	
YLL		4.65	5.06	4.63	YIR ^b	4.88	4.57	4.7	YYK	1	1.43	1.1	
YVL ^a		4.77	6.09	4.77	YFH	3.18	3.13	2.99	FFK	2.48	2.67	2.39	
YIV		5.3	5.22	5.4	YIL ^a	5.65	5.31	5.54	YIK	4.89	4.59	4.72	
ACE	FYK ^b	1	0.47	0.88	FIL	4.5	4.34	4.35	YFR ^{a,b}	4.3	3.82	5.67	
	VW ^{a,b}	5.8	5.26	5.93	IG ^{a,b}	2.92	2.62	2.68	GG ^{a,b}	2.14	2.01	1.91	
	IW	5.7	5.39	5.19	GI	2.92	2.72	2.69	QG	2.13	2.12	1.96	
	IY	5.43	4.99	5.53	GM	2.85	2.66	2.87	SG	2.07	1.95	1.88	
	AW	5	5.3	4.94	GA	2.7	2.65	2.65	LG	2.06	2.64	2.11	
	RW ^a	4.8	5.27	4.9	YG ^a	2.7	2.4	2.46	GD ^a	2.04	1.72	1.92	
	VY	4.66	4.86	4.66	GL	2.6	2.55	2.66	TG	2	2.13	2.1	
	GW ^b	4.52	4.78	4.42	AG ^b	2.6	2.54	2.43	EG ^b	2	1.84	1.77	
	VF	4.28	3.81	3.87	GH	2.51	2.58	2.7	DG	1.85	1.8	1.68	
	AY ^a	4.06	4.6	4.58	GR ^a	2.49	2.47	2.26	PG ^a	1.77	1.58	1.77	
	IP	3.89	3.78	3.75	KG	2.49	2.79	2.75	LA	3.51	3.28	3.45	
	RP ^b	3.74	3.67	3.45	FG ^b	2.43	2.36	2.48	KA ^b	3.42	3.43	3.49	
	AF	3.72	3.65	3.69	GS	2.42	2.35	2.24	RA	3.34	3.15	3.13	
	GY ^a	3.68	4.17	3.95	GV ^a	2.34	2.69	2.34	YA ^a	3.34	3.04	3.21	
	AP	3.64	3.69	3.5	MG	2.32	2.42	2.44	AA	3.21	3.18	3.17	
	RF	3.64	3.62	3.65	GK	2.27	2.58	2.47	FR	3.04	2.82	2.83	
	VP	3.38	3.65	3.58	GE	2.27	2.29	2.29	HL	2.49	2.56	2.81	
	GP ^{a,b}	3.35	3.17	2.97	GT ^{a,b}	2.24	2.59	2.44	DA ^{a,b}	2.42	2.44	2.42	
GF	3.2	3.13	3.17	WG	2.23	2.43	2.54	EA	2	2.48	2.31		
IF	3.03	3.74	3.44	HG	2.2	2.02	2.06						
VG	2.96	2.5	2.51	GQ	2.15	2.21	2.13						

^a Dataset I.
^b Dataset II.

Abbreviations: Obsd — observed value;
Pred — predicted value.

$$k = \sum(Y_{\text{obs}} \times Y_{\text{pred}}) / \sum(Y_{\text{pred}})^2 \quad (0.85 \leq k), \tag{7}$$

$$k' = \sum(Y_{\text{obs}} \times Y_{\text{pred}}) / \sum(Y_{\text{obs}})^2 \quad (k' \leq 1.15). \tag{8}$$

Here, Y_{obs} and Y_{pred} indicate the observed and predicted activities, while \bar{Y}_{obs} , and \bar{Y}_{pred} refer to the mean values of the observed and predicted activities, respectively. $\bar{Y}_{\text{training}}$ indicates the mean activity value of the training set. The squared correlation coefficient between the observed (Y axis) and predicted (X axis) values with the intercept (r^2) and without the intercept (r_0^2) are calculated to determine r_m^2 . A change in the axes gives the value of r'^2 . The k and k' parameters indicate the slopes in the former and latter cases, respectively.

The advantage of this metrics is that it is comparable to all other commonly used validation metrics in terms of the threshold value. The parameters were calculated at a web application (accessible from <http://aptsoftware.co.in/DTCMLRWeb/index.jsp>).

Splitting of the dataset. In order to prove the validity and stability of the model, the splitting of the dataset plays a crucial role in the development of a QSAR model. The dataset was divided into training and test sets [10] used for the model development and validation, respectively. The datasets were divided into training and test sets two times: in dataset I, one of every four samples was chosen as the test set (highlighted with 'a'), while the 1st, the 7th, the 11th, the 17th sample, etc. were chosen as the test set (highlighted with 'b') in dataset II.

RESULTS AND DISCUSSION

QSAR model for BTT. The results of the r_m^2 parameter in dataset I obtained for each model are given in Fig. 1, *a*. It shows that r_m^2 changes with SMR-introduced variables, reaching the maximum at the 11th step. With r_m^2 up to 0.861, the optimal model (Eq. 9) was then selected. Likewise, the optimal model (Eq. 10) was selected when r_m^2 reached 0.870 (Fig. 1, *b*) at the 7th step in dataset II. All other calculated parameters are quite good and the satisfactory conditions are summarized in Table 2.

$$pT = 2.301 - 1.334E_{1-2}^{\text{1st}} - 0.356E_{1-1}^{\text{2nd}} + 1.098E_{1-8}^{\text{2nd}} - 0.001H_{1-3}^{\text{1st}} - 15.461S_{1-1}^{\text{1st}} - 1.132E_{1-5}^{\text{2nd}}, \quad (9)$$

$$pT = 0.860 - 0.493E_{1-3}^{\text{1st}} + 1.014H_{1-9}^{\text{1st}} - 0.142E_{2-8}^{\text{1st}}. \quad (10)$$

Based on Eq. 9, we consider the independent variable E_{1-8}^{2nd} to be in the positive correlation with the dependent variable pT , while the independent variables E_{1-2}^{1st} , E_{1-1}^{2nd} , H_{1-3}^{1st} , S_{1-1}^{1st} , and E_{1-5}^{2nd} are in negative correlations with pT . (Since some interaction items have a zero coefficient, they are not listed in the equation. The superscript of variables denotes the position for the interaction (the same below)). It can be seen that the H—H steric interaction in the 1st amino acid is the most important factor affecting pT . Eq. 10 shows that the H— O_{sp^2} electrostatic interaction in the 1st amino acid is the key factor affecting pT . Variation of the key factors can greatly improve pT . Fig. 2, *a* (I) and 2, *b* (II) presents the scatter plot of the observed values versus the predicted ones (Table 1) of BTT. The solid line indicates the fitted line of the training set and the dotted line indicates the fitted line of the test set (the same below). The activities predicted using Eq. 9 were plotted against the observed activity data, and the

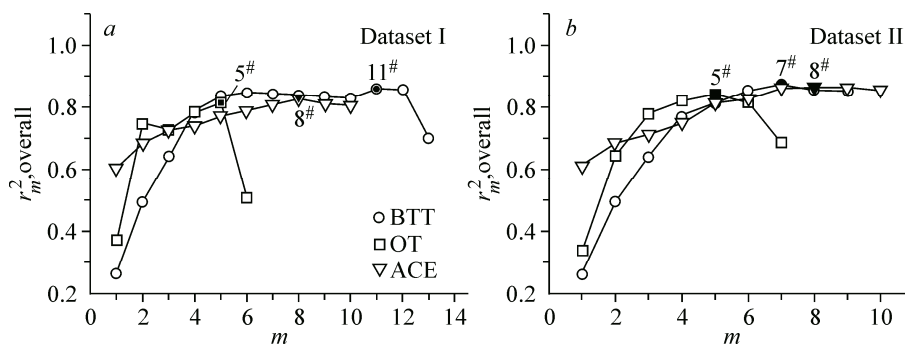


Fig. 1. Comparison of r_m^2 of different models in dataset I and dataset II

Table 2

Results of the different validation metrics for the MLR model and the satisfactory conditions

Criteria	Calculated value						Satisfactory value
	BTT		OT		ACE		
Q_{LOO}^2	0.903 ^a	0.879 ^b	0.897 ^a	0.970 ^b	0.883 ^a	0.852 ^b	> 0.5
r^2	0.875 ^a	0.939 ^b	0.872 ^a	0.955 ^b	0.891 ^a	0.955 ^b	> 0.6
r_0^2	0.861 ^a	0.937 ^b	0.855 ^a	0.777 ^b	0.871 ^a	0.949 ^b	$(r_0^2 \text{ or } r_0'^2) \approx r^2$
$r_0'^2$	0.874 ^a	0.939 ^b	0.872 ^a	0.881 ^b	0.890 ^a	0.938 ^b	
$(r^2 - r_0^2)/r^2$	0.014 ^a	0.00189 ^b	0.019 ^a	0.187 ^b	0.022 ^a	0.006 ^b	one of these criteria < 0.1
$(r^2 - r_0'^2)/r^2$	0.0002 ^a	0.00031 ^b	0.00003 ^a	0.078 ^b	0.001 ^a	0.017 ^b	
K	0.984 ^a	0.993 ^b	0.960 ^a	1.103 ^b	0.972 ^a	1.060 ^b	$0.85 \leq k \text{ or } k' \leq 1.15$
k'	1.006 ^a	1.00102 ^b	1.015 ^a	0.883 ^b	1.014 ^a	0.938 ^b	
r_m^2	0.861 ^a	0.870 ^b	0.811 ^a	0.841 ^b	0.826 ^a	0.865 ^b	> 0.5
$r_m'^2$	0.852 ^a	0.820 ^b	0.884 ^a	0.887 ^b	0.836 ^a	0.823 ^b	

^a Dataset I.^b Dataset II.

resulting graph (Fig. 2, *a*) showed that the points were minutely scattered about the line of fit, i.e., the results are relatively close to the observed values and the model possesses the strong fitting capacity, and so does Fig. 2, *b* (II). Therefore, the models may be considered satisfactory for predicting the activity of a new set of such peptide drugs.

QSAR model for oxytocin. The scales combined with SMR to select the variables, and then a MLR model with the best optimal number of variables is built. In Fig. 1, *a*, it was shown that r_m^2 changed with SMR-introduced variables, reaching the maximum (0.811) at the 5th step in dataset I; the optimal model is Eq. 11. The optimal model (Eq. 12) was selected when r_m^2 reached 0.841 (Fig. 1, *b*) at the 5th step in dataset II. The other calculated parameters summarized in Table 2 are fairly satisfactory.

$$OA = -78.874 - 18.948E_{3-8}^{2\text{nd}} + 0.196H_{5-9}^{1\text{st}} - 60.167E_{5-9}^{2\text{nd}} + 3.684H_{8-9}^{1\text{st}}, \quad (11)$$

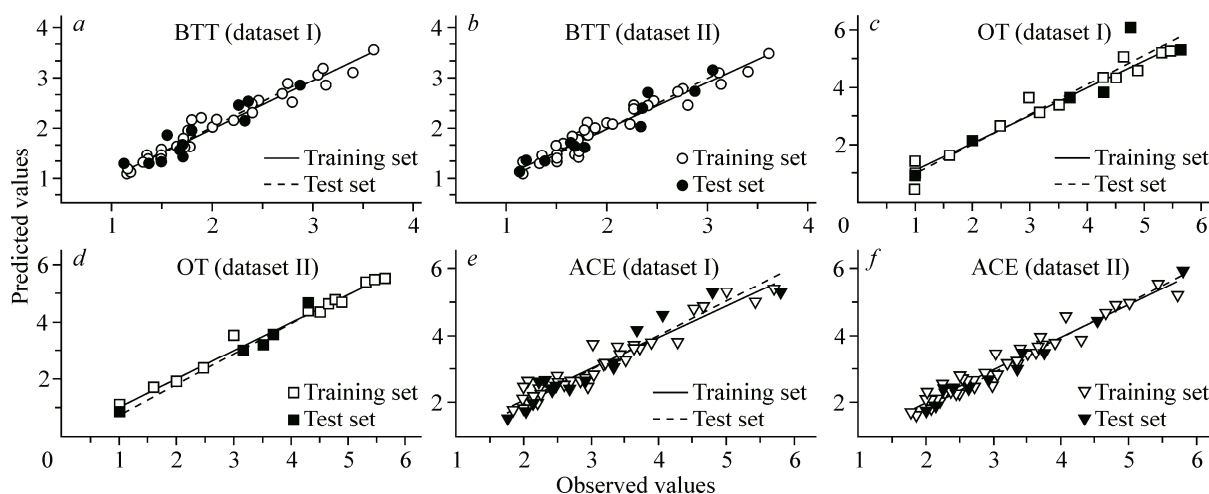


Fig. 2. Scatter plot of the observed vs. predicted values for the MLR model developed of BTT, OT, and ACE

$$OA = -27.198 + 0.013H_{3-9}^{1st} + 3.509E_{1-8}^{2nd}. \quad (12)$$

In accordance with Eq. 11, it can be seen that OA is influenced by two interactions: electrostatic and hydrophobic interactions. The coefficients of E_{3-8}^{2nd} , E_{5-9}^{2nd} , H_{8-9}^{1st} variables are relatively large, and this shows that $C_{sp2}-O_{sp3}$ electrostatic interactions in the 2nd amino acid, $N_{sp3}-O_{sp2}$ electrostatic interactions in the 2nd amino acid, and $O_{sp3}-O_{sp2}$ hydrophobic interactions in the 1st amino acid are important factors affecting OA , especially the former two interactions. According to the coefficients of model (Eq. 12) variables, the $H-O_{sp3}$ electrostatic interaction in the 2nd amino acid has a significant impact. The pT value can be substantially improved by varying the important interactions. Fig. 2, *c* (I) and 2, *d* (II) show that the results in both are relatively close to the observed values and uniformly dispersed around the fitted line, with the exception of 5[#] in Fig. 2, *c*. The errors can be caused by the characterization method, experiment, or other reasons. Whatever the reasons, there was only one sample with large errors, and the excellent results of this study should be recognized. Thus, the predictive ability of the model was found to be satisfactory.

QSAR model for angiotensin converting enzyme. The optimal model (Eq. 13 (I)) was selected when r_m^2 reached 0.826 (Fig. 1, *a*), while the optimal model (Eq. 14) was selected when r_m^2 reached 0.865 (Fig. 1, *b*) at the 8th step in dataset II. The other calculated parameters are summarized in Table 2.

$$pIC_{50} = 2.013 + 47.655S_{1-1}^{1st} - 7.466E_{3-3}^{2nd} - 0.511E_{1-3}^{1st} + 0.005H_{1-9}^{2nd} - 3.173E_{2-3}^{2nd}, \quad (13)$$

$$pIC_{50} = 1.732 + 52.148S_{1-1}^{1st} - 0.604E_{1-3}^{1st} - 5.824E_{3-3}^{2nd} + 0.010S_{2-2}^{2nd} + 0.022S_{2-8}^{2nd} - 34.914S_{1-1}^{2nd} - 0.802E_{1-2}^{2nd}. \quad (14)$$

It is seen from Eq. 13 that the coefficients of S_{1-1}^{1st} , E_{3-3}^{2nd} , and E_{2-3}^{2nd} variables are more than 1. Especially S_{1-1}^{1st} , the H—H steric interactions in the 1st amino acid, is the key factor affecting pIC_{50} . In Eq. 14 it is shown that pIC_{50} is greatly affected by S_{1-1}^{1st} (H—H steric interactions) and S_{1-1}^{2nd} . A plot of the predicted (Table 1) against observed values yielded a straight line with points well predicted about the line (Fig. 2, *e*) in dataset I. The pIC_{50} values of all the peptide sequences predicted using Eq. 14 were plotted against the observed values, and the resulting graph (Fig. 2, *f*) showed that the points were minutely scattered about the line of fit. The established models will help in the design of novel peptide drugs.

CONCLUSIONS

In this paper, the MLR model has been developed for different peptide drugs with 3D-HoVAIF. Its predictive ability and robustness are reflected by the r_m^2 parameters: leave-one-out (Q_{LOO}^2) for the internal validation and predictive (r^2) for the external validation. Satisfying results showed that the QSAR models constructed with 3D-HoVAIF had a good predictive ability and strong robustness. Hence, 3D-HoVAIF proves to be an effective methodology for the characterization of complex interactions of amino acid molecules and the prediction of bioactivity of peptide drugs. Moreover, it will provide theoretical guidance for the design of new peptide drugs.

This work was supported by the National Natural Science Funds of China (21475081) the Scientific Research Planning Program of Shaanxi Province of China (2015 JM 2057), and the Graduate Innovation Fund of Shaanxi University of Science and Technology.

REFERENCES

1. Jiráček J., Yiotakis A., Vincent B., Lecoq A., Nicolaou A., Checler F., Dive V. // *J. Biol. Chem.* – 1995. – **270**. – P. 21701 – 21706.
2. Marraud M., Aubry A. // *J. Pept. Sci.* – 1996. – **40**. – P. 45 – 83.
3. Sneath P.H.A. // *J. Theor. Biol.* – 1966. – **12**. – P. 157 – 195.
4. Kidera A., Konishi Y., Oka M., Ooi T., Scheraga H.A. // *J. Protein Chem.* – 1985. – **4**. – P. 23 – 55.
5. Zhou P., Zhou Y., Wu S.R., Li B., Tian F.F., Li Z.L. // *Chinese Sci. Bull.* – 2006. – **51**. – P. 524 – 529.
6. Hellberg S., Eriksson L., Jonsson J., Lindgren F., Sjoström M., Skagerberg B., Wold S., Andrews P. // *Int. J. Pept. Res. Ther.* – 1991. – **37**. – P. 414 – 424.

7. Hellberg S., Sjoestroem M., Skagerberg B., Wold S. // *J. Med. Chem.* – 1987. – **30**. – P. 1126 – 1135.
8. Hellberg S., Sjostrom M., Wold S. // *Acta Chem. Scand. B: Org. Chem. Biochem.* – 1986. – **40**. – P. 135 – 140.
9. Wold S., Eriksson L., Hellberg S., Jonsson J., Sjöström M., Skagerberg B., Wikström C. // *Can. J. Chem.* – 1987. – **65**. – P. 1814 – 1820.
10. Tong J.B., Chen Y., Liu S.L., Che T., Xu X.M. // *J. Chemom.* – 2012. – **26**. – P. 549 – 555.
11. Hemmateenejad B., Yousefinejad S., Mehdipour A.R. // *Amino Acids.* – 2011. – **40**. – P. 1169 – 1183.
12. Sandberg M., Eriksson L., Jonsson J., Sjöström M., Wold S. // *J. Med. Chem.* – 1998. – **41**. – P. 2481 – 2491.
13. Wang X.Y., Wang J., Lin Y., Ding Y., Wang Y.Q., Cheng X.M., Lin Z.H. // *J. Mol. Model.* – 2011. – **17**. – P. 1599 – 1606.
14. Tong J.B., Li Y.F., Liu S.L., Meng Y.L. // *Chin. J. Struct. Chem.* – 2010. – **12**. – P. 1893 – 1899.
15. Zhou P., Tian F.F., Li Z.L. // *Chemom. Intell. Lab. Syst.* – 2007. – **87**. – P. 88 – 94.
16. Collantes E.R., Dunn W.J. // *J. Med. Chem.* – 1995. – **38**. – P. 2705 – 2713.
17. Mei H., Zhou Y., Sun L.L., Li Z.L. // *Acta Phys. Chim. Sin.* – 2004. – **20**. – P. 821 – 825.
18. Cocchi M., Johansson E. // *Quant. Struct. – Act. Rel.* – 1993. – **12**. – P. 1 – 8.
19. Jain A., Zongker D. // *IEEE Trans. Pattern Anal.* – 1997. – **19**. – P. 153 – 158.
20. Guttman I. *Linear Models: an Introduction.* – New York: Wiley, 1982.
21. Morrison D.F. *Multivariate Statistical Methods*, 3rd ed. – New York: McGraw-Hill, 1990.
22. Roy K., Mitra I., Kar S., Ojha P.K., Das R.N., Kabir H. // *J. Chem. Inf. Model.* – 2012. – **52**. – P. 396 – 408.
23. Roy K., Chakraborty P., Mitra I., Ojha P.K., Kar S., Das R.N. // *J. Comput. Chem.* – 2013. – **34**. – P. 1071 – 1082.
24. Golbraikh A., Tropsha A. // *Mol. Diversity.* – 2000. – **5**. – P. 231 – 243.
25. Golbraikh A., Shen M., Xiao Z., Xiao Y.D., Lee K.H., Tropsha A. // *J. Comput. – Aided Mol. Des.* – 2003. – **17**. – P. 241 – 253.