Received: 15 May 2017 | Revised: 11 August 2017 | Accepted: 13 September 2017

DOI: 10.1111/cbdd.13150

RESEARCH ARTICLE

WILEY

Predictive QSAR modeling study on berberine derivatives with hypolipidemic activity

Pan Yu[1,2,†] | **Dongdong Li**[1,2,†] | **Junjun Ni**[2] | **Linguo Zhao**[1,2] | **Gang Ding**[3] | **Zhenzhong Wang**[3] | **Wei Xiao**[3]

[1]Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China

[2]College of Chemical Engineering, Nanjing Forestry University, Nanjing, China

[3]Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang, Jiangsu Province, China

Correspondence
Linguo Zhao, College of Chemical Engineering, Nanjing Forestry University, Nanjing, China.
Email: lg.zhao@163.com
and
Wei Xiao, Jiangsu Kanion Pharmaceutical Co., Ltd., Lianyungang, China.
Email: xw@kanion.com

Funding information
The Jiangsu "333" project of cultivation of high-level talents, Grant/Award Number: BRA2015317; the 11th Six Talents Peak Project of Jiangsu Province, Grant/Award Number: 2014-JY-011; A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD)

Berberine (BBR), isolated from a Chinese herb, is identified as a new cholesterol-lowering small molecule, and hundreds of berberine derivatives have been obtained for optimization of their hypolipidemic activities in recent years. However, so far there is no available quantitative structure–activity relationship (QSAR) model used for the development of novel BBR analogues with hypolipidemic activities, mainly due to lack of lipid-lowering molecular mechanisms and target identification of BBR. In this paper, the tactics using ligand efficiency indice instead of pIC_{50} as the activity could be adopted for the development of BBR QSAR models. A series of 59 BBR derivatives with hypolipidemic activities have been studied and split randomly into three sets of training and test sets. Statistical quality of most building models shows obviously robust. Best calculated model that employs LLE indice as the activity (Model **6**) has the following statistical parameters: for training set $R^2 = .984$, $Q^2 = 0.981$, RMSE = 0.1160, and for test set $R^2 = .989$, RMSE = 0.0067. This model would be used for the development of novel BBR analogues with lipid-lowering activities as a hit discovery tool.

KEYWORDS
berberine, hypolipidemic activity, ligand efficiency indice, molecular descriptors, QSAR

1 | INTRODUCTION

Berberine (BBR, Figure 1), an important isoquinoline alkaloid, is widely distributed in various medicinal herbs and has been used in traditional Chinese medicine for hundreds of years. Berberine displays various pharmacological activities that have applications in different therapeutic areas such as cancer, inflammation, diabetes, and cardiovascular diseases.[1,2] Especially in treating hyperlipidemia diseases, BBR has been identified as a new cholesterol-lowering drug with a mechanism of action different from that of statin drugs, mainly through a post-transcriptional mechanism that stabilizes the low-density lipoprotein receptor (LDLR) mRNA.[3] Nowadays, there are two detailed studies on how to stabilize the LDLR mRNA by BBR: One is involved with regulation of hepatocyte nuclear factor 1 (HNF1) which influences PCSK9 gene transcription,[4,5] and the other is associated with activating AMPK through mitochondria inhibition, and subsequently extracellular-signal-regulated kinase

Abbreviations:
AMPK, AMP-activated protein kinase; BBR, berberine; ERK, extracellular-signal-regulated kinase; HNF1, hepatocyte nuclear factor 1; LDLR, low-density lipoprotein receptor; LE, ligand efficiency; LOO, leave-one-out; PCSK9, proprotein convertase subtilisin/kexin type 9; QSAR, quantitative structure–activity relationship.

[†]These authors equally contributed to this work.

(ERK)-mediated LDLR mRNA up-regulation.[6,7] However, the detailed profile of lipid-lowering targets closely related to LDLR mRNA up-regulation has not been elucidated clearly yet.

Based on this, dozens of BBR analogues reported by Jiang's laboratory[8–11] have been designed, synthesized, and investigated for its cholesterol-lowering activity according to conventional medicinal chemistry campaign. The analysis of structure–activity relationship (SAR) (Figure 1) indicates that the methylenedioxy group at the 2- and 3-position is an essential element to retain the activity, and hydrophobic groups at the 9-position can increase the activity. To some extent, although this kind of SARscan guides the rational design of novel BBR-like lipid-lowering drug, its application domain actually seems very limited.

QSAR modeling is an important virtual screening tool capable of quantitatively predicting the bioactivity of the explored compounds.[12] Currently, there is no predictive QSAR/QSPR modeling successfully employed to investigate lipid-lowering biological activities and chemical properties of BBR analogues. The limited data curated from the corresponding references in quality and quantity hamper the development of the QSAR modeling of BBR analogues with hypolipidemic activity. Several papers[13–15] have described that a ligand efficiency (LE) indice instead of pIC_{50} is used as the activity in QSAR, and better fits and predictions can be obtained with ligand efficiency. Thus, it would be a good choice for BBR model building to employ suitable LE indice.

The aim of this study is to build QSAR models based on 59 BBR derivatives with hypolipidemic activity reported by Jiang's laboratory. In this study, we attempt to build QSAR models in which four common ligand efficiency indice (BEI, LLE, SEI, and LELP, see Section 2.2) instead of pIC_{50} could be used as the activity. The obtained results for the four types of LE indice show significant improvement over pIC_{50}. Surely this is strongly correlated with the computable physical property included in LE definition. Among these models, the best QSAR model was analyzed and discussed strictly.

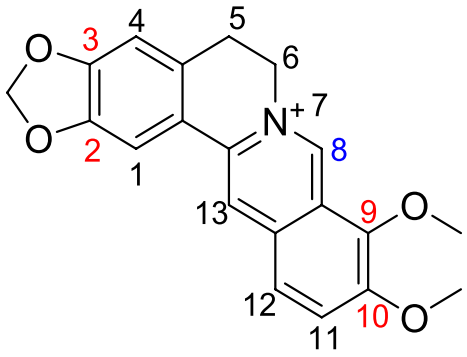

FIGURE 1 Chemical structure of berberine (BBR) [Colour figure can be viewed at wileyonlinelibrary.com]

2 | METHODS AND MATERIALS

2.1 | Data set and splitting

A series of 59 berberine derivatives (berberine included) from Jiang's laboratory[8–11] were used for the establishment of QSAR model. General structure and hypolipidemic activities of derivatives are presented in Table S1, which shows antilipidemic activity characterized with the expression level of LDLR mRNA. Due to the different measuring concentrations of compounds **18**–**34** abstracted from the literature,[9] activity values of these compounds have been adjusted. In addition, antilipidemic activities of all the compounds were converted to their reciprocals T for subsequent data pretreatment.

Data set splitting plays a key role for the successful development of a QSAR model. All the compounds were divided into three random splits (20% used in test set while 80% in training set). All the three splits should satisfy the following three principles[12]: (i) the range of the activity is roughly the same for each subclass; (ii) the splits are random; (iii) the splits are not identical (Table S2). We have checked that level of identity between these splits and these random splits is different enough (data not shown).

2.2 | Definitions of LE

According to Robert's study,[13] we have chosen four LE indice as QSAR Activities: BEI, LLE, SEI, and LELP.

$$\mathrm{BEI} = \mathrm{pActivity}/\mathrm{NA} \tag{1}$$

where NA is the number of non-hydrogen atoms within molecules, while the value of pActivity is calculated by the same form with pIC_{50}.

$$\mathrm{LLE} = \mathrm{pActivity} - \mathrm{ALOGP} \tag{2}$$

where ALOGP, the important part for LLE, is calculated by the Canvas module of the Schrodinger 2015 suite.

$$\mathrm{SEI} = 100 \times \mathrm{pActivity}/\mathrm{TPSA} \tag{3}$$

where TPSA, the key property of SEI, is the solvent accessible polar surface area and is calculated by Discovery studio 3.5. It is obviously noted that TPSA is closely involved with polar atoms within molecules. This makes it hard to make predictive models based on SEI for molecules that contain many nonpolar atoms. To circumvent this, molecule with TPSA <20 $Å^2$ is removed.

$$\mathrm{LELP} = \mathrm{CLOGP}/\mathrm{BEI} = \mathrm{CLOGP} \times \mathrm{NA}/\mathrm{pActivity} \tag{4}$$

where CLOGP comes from Sybyl-X 2.1. All the above are calculated and retained in Table S3.

2.3 | Molecular descriptors calculation and selection

Three different sources of molecular descriptors (a set of 57 kinds of 3D descriptors from Discovery studio 3.5, a set of 237 kinds of physicochemical descriptors and topological descriptors from the Canvas module of Schrodinger 2015 suite, and a set of 340 kinds of 2D/3D descriptors from MOE2014) are selected and stepwise calculated by means of each corresponding software above.

Molecular descriptors used for the establishment of QSAR model are selected by running QuaSAR-Contingency in MOE2014. QuaSAR-Contingency is a statistical application in the selection of descriptors for QSAR or QSPR, which can perform a bivariate contingency analysis for each descriptor and the activity or property value. It produces a table of coefficients that assist to select important descriptors. Three coefficients (*B*, *C*, and *V*) in contingency analysis can be regarded as the measures of association.

$$B = \sum_{i,j} (m_{i,j} - n_{i,j}) \tag{5}$$

where *i* and *j* range over a collection of bins, or intervals, into which the samples are deposited.

$$C = \sqrt{B/(B+m)} \tag{6}$$

$$V = \sqrt{B/(m \times \max(I-1, J-1))} \tag{7}$$

where *I* and *J* are the number of *i* and *j* bins, respectively. All the measures are in the range [0, 1], where 1 means perfect dependence and 0, no dependence. Definitions of significant descriptors are shown in Table S4.

2.4 | QSAR methods

We would determine the parameters of the linear model by the method of partial least squares (PLS) in MOE2014. The number of components (up to 10) is optimized by cross-validation on the training set. All the parameters of the models could be displayed in Table S4.

2.5 | Validation of QSAR model

Three common validation methods are applied to evaluate the predictive performance of the QSAR models: cross-validation using the training set, external validation using the test set, and data randomization.

Leave-one-out (LOO) cross-validation is one of the classical techniques used to developed models as an internal validation. The two parameters cross-validated Q^2 and root-mean-square error (RMSE) determine the predictive ability of the model.[16] Higher value of Q^2 and lower value of RMSE display better model prediction.

$$Q^2 = 1 - \sum (Y_{\text{obs}} - Y_{\text{pred}})^2 / \sum (Y_{\text{obs}} - \bar{Y}_{\text{train}})^2 \tag{8}$$

$$\text{RMSE} = \sqrt{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2 / n} \tag{9}$$

In Equations 8 and 9, Y_{obs} is the observed property of the training set compounds, Y_{pred} is the LOO-predicted property of the training set compounds, Y_{train} is the mean observed property of the training set compounds, and *n* refers to the number of observations. The predictive ability of model is considered as acceptable when Q^2 is greater than 0.5.

Novel statistical metric (R^2_{m}) was used to validate a true predictive potential of developed QSPR models.[17,18] For calculating R^2_{m} metric, an open-access Web application "Rmsquare Calculator" is available at http://aptsoftware.co.in/rmsquare/.

Y-randomization test was frequently used for checking the robustness of the QSAR model. For a reliable QSPR model, the average correlation coefficient (R_{r}) of randomized models should be less than the correlation coefficient (R) of non-randomized model. A parameter CR^2_{p} penalizes the model R^2 for a small difference between squared mean correlation coefficient (R^2_{r}) of randomized models and squared correlation coefficient (R^2) of the non-randomized model. CR^2_{p} can be defined as:

$$\text{CR}^2_{\text{p}} = R \times (R^2 - R^2_{\text{r}})^2 \tag{10}$$

The value of CR^2_{p} should be greater than 0.5 if an acceptable QSPR model.

3 | RESULTS

A total of 58 BBR analogues with hypolipidemic activity plus BBR as the data set were carefully collected from four papers published by Jiang's laboratory.[8–11] It was known that the activity values for these compounds were represented by up-regulated levels of LDLR. Therefore, the relative values of these BBR analogues can be dealt by reciprocal transformation presented in Table S1. It can be seen that several compounds with different substituents could exhibit the same activity values. However, the suitable QSAR models cannot

TABLE 1 Statistical quality parameters of QSAR models of berberine derivatives with hypolipidemic activity for three random splits

Split	Model	Ligand efficiency indice	Number of descriptors	R^2_{train}	$RMSE_{train}$	F_{train}	R^2_{LOO}	$RMSE_{LOO}$	R^2_{test}	$RMSE_{test}$	F_{test}	$R^2_{m(train)}$
Split 1	1	BEI	4	0.947	0.0056	881	0.934	0.0064	0.970	0.0045	1,177	0.9478
	2	LLE	3	0.982	0.1121	2,676	0.979	0.1220	0.990	0.0817	4,080	0.9828
	3	SEI	4	0.813	2.5598	246	0.712	3.2189	0.952	1.3674	766	0.8130
	4	LELP	5	0.712	3.3123	160	0.565	4.1073	0.468	4.9381	70	0.7122
Split 2	5	BEI	6	0.940	0.0064	769	0.908	0.0080	0.706	0.0072	144	0.9402
	6	LLE	4	0.984	0.1160	2,899	0.981	0.1258	0.989	0.0067	2,706	0.9828
	7	SEI	4	0.835	2.5230	278	0.709	3.3843	0.842	1.6316	275	-[a]
	8	LELP	6	0.727	3.1440	168	0.654	3.543	0.267	6.3224	56	0.7268
Split 3	9	BEI	4	0.942	0.0056	794	0.934	0.0059	0.985	0.0064	781	0.9421
	10	LLE	3	0.980	0.1139	2,300	0.977	0.1225	0.993	0.1215	3,751	0.9800
	11	SEI	3	0.820	2.0778	255	0.760	2.4046	0.845	3.4403	296	0.8199
	12	LELP	7	0.748	3.0300	183	0.661	3.5239	0.848	3.3426	195	-[a]

[a]No result.

be obtained by the treated value as the activity in QSAR (data not shown).

From data set presented in Table S2, three random splits were generated. Generated training and test sets had 48 and 11 molecules, respectively (20% of compounds were used in test set). The three splits are different as the level of identity between them should be kept low. The search for the best model included models with four LE indice in three splits. Moreover, the values of four LE indice of all the compounds in Table S3 have been calculated as detailed in Section 2.2. There are 12 new QSAR models built based on the 3–7 kinds of molecular descriptors related to LE indice. The detailed parameters and corresponding linear models of these QSAR models are displayed in Table S4. Three categories of Models 1–4, 5–8, and 9–12 are established using the data set in Split 1, Split 2, and Split 3, respectively. As to descriptors selection, the QuaSAR-Contingency in MOE2014 is employed for the selection process. The two coefficients (*B* and *V*) in contingency analysis could determine the types and quantity of molecular descriptors.

4 | DISCUSSION

Table 1 shows the statistical quality of the built models. The presented data display that statistical quality for calculated models from Split 1 to 3 possesses the same trends. All the models exhibit good predictive potency, which suggest that ligand efficiency indice instead of pIC_{50} used as the activity in QSAR seems apparent superiority. Results also indicate that two LEs (BEI and LIE) are superior to the other two LEs in the development of QSAR models. It is known that LE is strongly involved (positively or negatively) with the computable physical property included in its definition, and the physical property related with the ligand efficiency is easier to predict than pIC_{50} in QSAR model.[13] Based on this, it can be observed in Table S4 that several molecular descriptors such as VDistMa, AlogP, chi1v_C employed in the BEI or LIE QSAR models (Models 1, 2, 5, 6, 9, and 10) correlate closely with BEI or LIE definition. The best calculated model was Model 6 for Split 2. Moreover, the QSAR models used by LIE ligand efficiency indice (Models 2, 6, 10) exhibit better statistical quality than others. Figure 2 describes best models for all splits graphically where highest value for R^2_{train} and R^2_{test} for the best model of each split is represented.

Statistical criteria of the predictability of the top six QSAR models for these three splits are displayed in Table 2.[16–18] The obtained results show that the predictability for the top six models seems very good. The value of R^2_{test} in most models except Model 5 is not less than 0.97, which reveals that most of these models have promising predictive potency. Among these models, the best predictive model is also Model 6, in

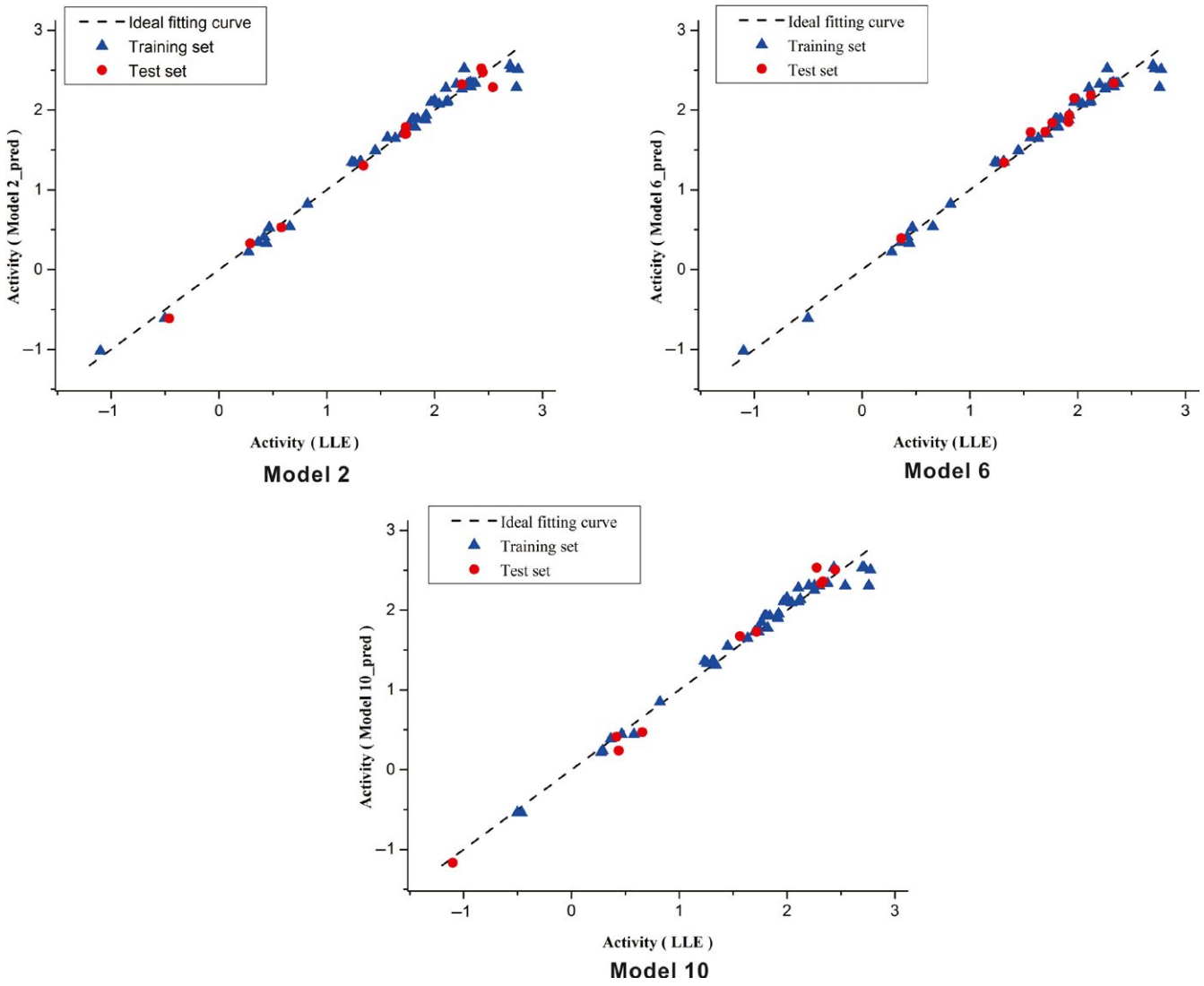

FIGURE 2 Graphical representation of QSAR models for best model in each split [Colour figure can be viewed at wileyonlinelibrary.com]

which R^2_{test} reaches to 0.989. In addition, all the six models for LE indice are satisfactory from the point of view of new criteria suggested by Roy et al.[17]

Y-randomization[19] is another tool widely used in validation of QSAR models. Table 3 shows Y-randomization for the top six QSAR models of berberine derivatives, which also confirms the robustness of suggested models. Consistent with the previous validation, Model 6 is the best predictive model of these six models. The corresponding value of CR^2_p is 0.9786 for train test and 0.9705 for test set, respectively. Four kinds of molecular descriptors (AlogP, chi1v_C, chi1_C, and logP (o/w)) are selected for the establishment of the QSAR Model 6 by the method of partial least squares (PLS). The detailed parameter in the estimated linear model was according to Equation 11:

$$\begin{aligned} LLE = {} & 6.10180 - 0.79515 \times AlogP - 0.32108 \times chilv\_C \\ & + 0.16998 \times chil\_C - 0.08273 \times logP(o/w) \end{aligned} \quad (11)$$

(training set: $R^2 = .984$, $Q^2 = 0.981$, RMSE = 0.1160, and test set: $R^2 = .989$, RMSE = 0.0067)

Ligand efficiency indice can be used as activities in QSAR models, and there are obvious improvement at the validation and prediction phase of QSAR models, compared with the common pIC_{50} as the activity.[13] The two factors could explain all the variations in the improvement of predictivity among data sets: One is the correlation between LE and physical property in its definition, and the other is the simplicity of physical property compared with pIC_{50}. Results presented in this work reveal that molecular descriptors used for construction of QSAR models are correlated closely with physical property defined within ligand efficiency indice. It could be surmised that this is the reason why the LE indice exhibit better than pIC_{50} as the activity in QSAR problems. Based on the current OECD guidelines,[20] high Q^2 cannot be a main parameter to ensure the predictive ability of a QSAR

TABLE 2 Criteria of predictability for the top six QSAR models of berberine derivatives

	Model 1	Model 2	Model 5	Model 6	Model 9	Model 10
$R^2_{test}(x,y)$: y-experimental values; x-calculated values						
R^2_{test}	0.970	0.990	0.706	0.989	0.985	0.993
R^2_0	0.9990	0.99996	0.9795	0.9942	0.99998	0.9978
R'^2_0	0.9985	0.99985	0.9778	0.9946	0.99988	0.9981
R_A	0.0299	0.01006	0.3874	0.0052	0.0152	0.0049
R_B	0.0294	0.00995	0.3850	0.0057	0.0151	0.0051
k	1.0034	1.0033	1.0076	0.9721	1.0005	0.9703
k'	0.9962	0.9935	0.9915	1.0271	0.9987	1.0268
$R^2_{m(test)}$	0.8831	0.9511	0.6445	0.9795	0.7794	0.9403
$R^2_{test}(x,y)$: x-experimental values; y-calculated values						
R^2_{test}	0.970	0.990	0.706	0.989	0.985	0.993
R^2_0	0.9985	0.99985	0.9778	0.9946	0.99988	0.9981
R'^2_0	0.9990	0.99996	0.9795	0.9942	0.99998	0.9978
R_A	0.0294	0.00995	0.3850	0.0057	0.0151	0.0051
R_B	0.0299	0.01006	0.3874	0.0052	0.0152	0.0049
k	0.9962	0.9935	0.9915	1.0271	0.9987	1.0268
k'	1.0034	1.0033	1.0076	0.9721	1.0005	0.9703
$R^2_{m(test)}$	0.917193	0.9602	0.5392	0.9692	0.8273	0.9476
$R^2_{m(av)}$	0.900168	0.9557	0.5919	0.9744	0.8033	0.9439
ΔR^2_m	0.034048	0.0091	0.1053	0.0103	0.0479	0.0073

$R_A = (R^2 - R^2_0)/R^2$ should be <0.1
$R_A = (R^2 - R'^2_0)/R^2$ should be <0.1
Should be $1.15 > k > 0.85$ and $1.15 > k' > 0.85$.
$R^2_{m(test)}$: R^2_m of test set and should be >0.5; $R^2_{m(av)}$: average value of R^2_m and should be >0.5
ΔR^2_m should be <0.2.

TABLE 3 Y-randomization for the top six QSAR models of berberine derivatives

	Model 1		Model 2		Model 5		Model 6		Model 9		Model 10	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
0	0.9586	0.9990	0.9891	0.9882	0.9429	0.9955	0.9852	1	0.945	0.9968	0.9814	0.9954
1	0.0003	0.0032	0.0502	0.02	0.003	0.0185	0.0072	0.0913	0.0178	0.0431	0.0424	0.3451
2	0.0037	0.0346	0.0658	0	0.0065	0.0613	0.0095	0.0357	0.0061	0.056	0.0009	0.0516
3	0.0072	0.0482	0.0742	0.0194	0.0051	0.0495	0.0157	0.0287	0.0029	0.0038	0.0091	0.0351
4	0.0001	0.006	0.1189	0.0184	0.0007	0.0201	0.0005	0.0318	0.0074	0.0189	0.0021	0.0074
5	0.0151	0.31	0.0089	0.2272	0.0241	0.3237	0.0897	0.0672	0.1113	0.0083	0.0115	0.0839
6	0.0892	0.0821	0.0019	0.004	0.0159	0.1156	0.0399	0.0301	0.0063	0.0086	0.0143	0.0224
7	0.0191	0.0191	0.0156	0.1813	0.0062	0.308	0.0129	0.3361	0.0023	0.3048	0.0188	0.0878
8	0.0013	0.0901	0.0077	0.0435	0.0004	0.0451	0.0079	0.0364	0.0004	0.1439	0.0004	0.1208
9	0.0021	0.0334	0.0001	0.0412	0.0101	0.1575	0.0003	0.0109	0.0012	0.019	0.0631	0.0001
10	0.0002	0.0266	0.0033	0.1993	0.0019	0.1144	0.0103	0.0707	0.0361	0.1845	0.01	0.0009
R^2_r	0.0138	0.0653	0.0347	0.0754	0.0074	0.1214	0.0194	0.0739	0.0192	0.0791	0.0173	0.0755
CR^2_p	0.9552	0.9753	0.9775	0.9627	0.9401	0.9443	0.9786	0.9705	0.9394	0.9694	0.9752	0.9706

$CR^2_p = R \times (R^2 - R^2_r$; CR^2_p should be >0.5.

model. Therefore, two other validation methods (R^2_m and Y-randomization test) are used. Results indicated that presented approach is a powerful alternative to more popular QSAR methods.

5 | CONCLUSIONS

QSAR models for BBR derivatives with hypolipidemic activity have been built step by step. Four ligand efficiency indice as the activity in QSAR study incorporated with the PLS method provided by the MOE 2014 software are able to be an efficient tool to build up robust models with good statistical quality. The predictive potency of the applied approach has been tested with three random splits into the training and test sets. The robustness of model has been proven with LOO cross-validation, novel statistical metric (R^2_m), and Y-randomization test. Four kinds of molecular descriptors (AlogP, chi1v_C, chi1_C, and logP (o/w)) selected in best model are closely related to physical property in LLE indice's definition. Best calculated model with LLE indice (Model 6) has following statistical parameters: for training set $R^2 = .984$, $Q^2 = 0.981$, RMSE = 0.1160, and for test set $R^2 = .989$, RMSE = 0.0067. It can be used for the development of BBR analogues with lipid-lowering activity as a hit discovery tool.

ACKNOWLEDGMENTS

This work was supported by the Jiangsu "333" project of cultivation of high-level talents (Grant No. BRA2015317), the 11th Six Talents Peak Project of Jiangsu Province (Grant No. 2014-JY-011), and A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ORCID

Linguo Zhao http://orcid.org/0000-0001-5552-5750

REFERENCES

[1] M. Imenshahidi, H. Hosseinzadeh, *Phytother. Res.* **2016**, *30*, 1745.
[2] A. F. G. Cicero, A. Baggioni, *Adv. Exp. Med. Biol.* **2016**, *928*, 27.
[3] W. J. Kong, J. Wei, P. Abidi, M. H. Lin, S. Inaba, C. Li, Y. L. Wang, Z. Z. Wang, S. Y. Si, H. N. Pan, S. K. Wang, J. D. Wu, Y. Wang, Z. R. Li, J. W. Liu, J. D. Jiang, *Nat. Med.* **2004**, *10*, 1344.
[4] B. Dong, H. Li, A. B. Singh, A. Cao, J. Liu, *J. Biol. Chem.* **2015**, *290*, 4047.
[5] H. Li, B. Dong, S. W. Park, H. S. Lee, W. Chen, J. Liu, *J. Biol. Chem.* **2009**, *284*, 28885.
[6] W. S. Kim, Y. S. Lee, S. H. Cha, H. W. Jeong, S. S. Choe, M. R. Lee, G. T. Oh, H. S. Park, K. U. Lee, M. D. Lane, J. B. Kim, *Am. J. Physiol. Endocrinol. Metab.* **2009**, *296*, E812.
[7] Z. Li, J. Jiang, W. Kong, *Biol. Pharm. Bull.* **2014**, *37*, 1766.
[8] Y. Wang, Y. Wang, H. Zhang, W. Kong, Y. Li, F. Liu, R. Gao, T. Liu, J. Jiang, D. Song, *Bioorg. Med. Chem. Lett.* **2009**, *19*, 6004.
[9] P. Yang, D. Song, Y. Li, W. Kong, Y. Wang, L. Gao, S. Liu, R. Cao, J. Jiang, *Bioorg. Med. Chem. Lett.* **2008**, *18*, 4675.
[10] Y. Li, P. Yang, W. Kong, Y. Wang, C. Hu, Z. Zuo, Y. Wang, H. Gao, L. Gao, Y. Feng, N. Du, Y. Liu, D. Song, J. Jiang, *J. Med. Chem.* **2009**, *52*, 492.
[11] Y. Wang, W. Kong, Y. Li, S. Tang, Z. Li, Y. Li, Y. Shan, C. Bi, J. Jiang, D. Song, *Bioorg. Med. Chem.* **2012**, *20*, 6552.
[12] P. Gramatica, *QSAR Comb. Sci.* **2007**, *26*, 694.
[13] R. P. Sheridan, *J. Chem. Inf. Model.* **2016**, *56*, 2253.
[14] J. Li, F. Bai, H. Liu, P. Gramatica, *Chem. Biol. Drug Des.* **2015**, *86*, 1501.
[15] N. Sugaya, *J. Chem. Inf. Model.* **2014**, *54*, 2751.
[16] A. Golbraikh, A. Tropsha, *J. Mol. Graph. Model.* **2002**, *20*, 269.
[17] P. P. Roy, K. Roy, *Chem. Biol. Drug Des.* **2009**, *73*, 442.
[18] P. K. Ojha, I. Mitra, R. N. Das, K. Roy, *Chemometrics Intell. Lab. Syst.* **2011**, *107*, 194.
[19] P. K. Ojha, K. Roy, *Chemometrics Intell. Lab. Syst.* **2011**, *109*, 146.
[20] O. In, *PARIS* **2007**, *69*, 154.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Yu P, Li D, Ni J, et al. Predictive QSAR modeling study on berberine derivatives with hypolipidemic activity. *Chem Biol Drug Des*. 2018;91:867–873. https://doi.org/10.1111/cbdd.13150