Cite this: *Med. Chem. Commun.*,
2017, 8, 1835

Integrating docking scores and key interaction profiles to improve the accuracy of molecular docking: towards novel B-Raf^{V600E} inhibitors†‡

Chun-Qi Hu,^{abd} Kang Li,^{ac} Ting-Ting Yao,^a Yong-Zhou Hu,^a
Hua-Zhou Ying^{*a} and Xiao-Wu Dong ^{*a}

A set of ninety-eight B-Raf^{V600E} inhibitors was used for the development of a molecular docking based QSAR model using linear and non-linear regression models. The integration of docking scores and key interaction profiles significantly improved the accuracy of the QSAR models, providing reasonable statistical parameters ($R_{\text{train}}^2 = 0.935$, $R_{\text{test}}^2 = 0.728$ and $Q_{\text{CV}}^2 = 0.905$). The established MD-SVR (molecular docking based SMV regression) model as well as model screening of a natural product database was carried out and two natural products (quercetin and myricetin) with good prediction activities were biologically evaluated. Both compounds exhibited promising B-Raf^{V600E} inhibitory activities ($\text{IC}_{50}^{\text{Quercetin}} = 7.59 \mu\text{M}$ and $\text{IC}_{50}^{\text{Myricetin}} = 1.56 \mu\text{M}$), suggesting a high reliability and good applicability of the established MD-SVR model in the future development of B-Raf^{V600E} inhibitors with high efficacy.

Received 4th May 2017,
Accepted 14th July 2017

DOI: 10.1039/c7md00229g

rsc.li/medchemcomm

Introduction

In general, three isoforms in the Raf family exist, *i.e.* A-Raf, B-Raf, and C-Raf. The isoform B-Raf has been shown to be the major Raf effector in the MAPK signalling cascade and plays a significant role in sustained angiogenesis.¹ The frequency of B-RAF mutations varies widely in human cancers, with a particularly high frequency found in malignant melanoma (60–70%), thyroid (30–50%), ovarian (~30%) and colorectal carcinomas (5–20%).^{2,3} Among the detected activated mutations in B-Raf, B-Raf^{V600E}, a glutamic acid for valine substitution at residue 600, represents one of the most commonly mutated proto-oncogenes. Furthermore, the isoform plays a significant role in the development of numerous cancers of high clinical impact.⁴ Importantly, B-Raf^{V600E} represents a class of attractive targets for the discovery of small molecules with anticancer activity.⁵ To date, a series of oncogenic B-Raf^{V600E} inhibitors have been reported in various developmental stages.^{4,6} Among these spe-

cies, a variety of compounds have entered clinical trials in recent years, exhibiting promising clinical efficiencies.^{1,7} Notably, vemurafenib (PLX4032),⁸ a selective B-Raf^{V600E} inhibitor, has received FDA approval for the treatment of melanoma. However, some adverse dermatological reactions have been reported in the clinical use of vemurafenib,⁹ together with intrinsic resistances to current B-Raf inhibitors.¹⁰ As such, there is an urgent need for the discovery of novel structurally distinct small molecular inhibitors against B-Raf^{V600E} kinase with a safer administration profile.

Natural products exhibit an inherently large structural diversity and have been the major resource of bioactive agents in medical, nutritional and agrochemical research as well as life sciences. More specifically, the intrinsic complexity of natural product-based drug discovery has been integrated into various *in silico* screenings for the identification of potential hit compounds.¹¹ Recently, Shukla *et al.* have reported the identification of potential antileishmanial chemotherapeutic agents identified from a natural product database.¹² Wessjohann *et al.* have reported the discovery of novel androgens and anti-androgens from synthetic and natural product databases.¹³ All of these findings demonstrate the urgent rationale to identify novel compounds with different structural scaffolds, which are distinct from currently available inhibitors.¹⁴

There is now growing interest in applying mathematical methods to improve the accuracy of molecular docking, including multiple linear regression (MLR), partial least squares (PLS) and support vector machines (SVMs).¹⁵ In our previous study,¹⁶ we have demonstrated that these regression methods are indeed feasible to establish a reliable

^a Zhejiang Province Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, P.R. China.

E-mail: yybb@zju.edu.cn, dongxw@zju.edu.cn; Fax: +86 571 88981051;

Tel: +86 571 88981051

^b College of Chemistry & Chemical Engineering, Shaoxing University, Shaoxing, P. R. China

^c Collaborative Innovation Center of Yangtze River Delta Region Green Pharmaceuticals, Zhejiang University of Technology, Hangzhou, P.R. China

^d Yongning Pharma, Taizhou, P.R. China

† The authors declare no competing interests.

‡ Electronic supplementary information (ESI) available: The original data used to set up the models including Table S1 and the cross validation results Table S2. See DOI: 10.1039/c7md00229g

structure-based QSAR model by employing a ligand fit scoring function and the structural characteristics of the compounds. In doing so, we have been able to significantly improve the performance of QSAR analysis.¹⁷ As part of our continued interest in rational drug design,^{16,18–22} we herein describe the development and validation of linear as well as non-linear regression- and docking-based models for the computational identification of a series of B-Raf^{V600E} inhibitors. Three regression methods were applied to establish reliable, structure-based QSAR models. Lastly, the most promising SVM regression (SVR) model was employed in the virtual screening of novel B-Raf^{V600E} inhibitors. Furthermore, *in silico* virtual screening of a natural product database (NPD from the ZINC database) containing 90 000 commercially available compounds was performed. Through analysis of the screened results, two commercially available structures exhibited potentially useful activities (Scheme 1). The biological activities were then evaluated to validate the rationality of the screening.²³ The obtained results indicate that this MD-SVR (molecular docking based SVR) method features a promising prediction ability for the discovery of B-Raf^{V600E} inhibitors. The screened new compounds, potentially useful as B-Raf^{V600E} inhibitors, may be useful for further investigations and structural modifications in structure–activity relationship studies.

Computational method

Dataset

For molecular docking based QSAR modeling, a set of 98 molecules (Fig. 1) was selected from the literature.^{5,24–27} The corresponding B-Raf^{V600E} inhibitory activity data (Table 1) span over four orders of magnitude, from 0.18 nM to 5.4 mM. Biological data of B-Raf^{V600E} inhibitory activity [IC₅₀ (nM)] was converted to a negative log scale [–log IC₅₀] and was then used as a response variable (dependent variable) for the subsequent QSAR analysis. For model training and validation, the total dataset ($n = 98$) was randomly divided into a training set ($n = 79$) and a test set (n

= 19). The 3D-structures of all compounds were generated and optimized in Discovery Studio 2.5 software (Accelrys, Inc. San Diego, CA) using a CHARMM-like force field.²⁸ The experimental activities of pIC₅₀ of the training and test data as well as the relative error (RE) of the prediction model can be found listed in Table 2.

Molecular docking

Molecular docking of B-Raf^{V600E} inhibitors from the literature and a natural product database was performed using the LigandFit module in Discovery Studio 2.5.²⁹ Preparation for the target binding pocket: (a) the crystal structure of B-Raf^{V600E} (PDB:4EHG) was employed as the template for molecular docking with the LigandFit protocol;²⁹ (b) all crystallographic water molecules have been removed; (c) all hydrogen atoms were added using the CHARMM force field of the “Prepare Protein module”; (d) the “Find Sites as Volume of Selected Ligand” tool was applied to define the docking site of the resulting protein structure. Preparation for the ligand libraries: (a) variable numbers of Monte Carlo simulations were implemented for different conformer generation of ligands; (b) all atoms were added using the CHARMM force field; (c) all conformers were treated with the “ligand minimization” of the receptor–ligand interaction module.

LigandFit protocol: (a) the ligand–receptor interaction energies were calculated based on the Piecewise Linear Potential 1 (PLP1) force field;³⁰ (b) a short rigid body minimization was performed and the top pose for each ligand was saved; (c) scoring was performed with a set of scoring functions employed in the LigandFit module, including LigScore1_Dreiding, LigScore2_Dreiding, -PLP1, -PLP2, -PMF and Dock_Score.

Calculation and selection of descriptors

After performing the scoring function of the molecular descriptors, those that remain constant for all molecules were eliminated and variable pairs that have correlation coefficients greater than 0.75 were classified as pairs of variables that are associated with one of each associated pair. Then,

Scheme 1 The evolution of the establishment of PLS, MLR, and MD-SVR models and the application of MD-SVR in the database screening of novel B-Raf inhibitors. Computational method.

Med. Chem. Commun., 2017, 8, 1835–1844 | 1837

Table 1 Experimental and calculated BRAF^{V600} inhibitory activities by various methods (SVM, PLS and MLR) for compounds from the literature

Name	IC ₅₀ (nM)	log IC ₅₀	Prd. log IC ₅₀ by SVM	Prd. log IC ₅₀ by MLR	Prd. log IC ₅₀ by PLS	Name	IC ₅₀ (nM)	log IC ₅₀	Prd. log IC ₅₀ by SVM	Prd. log IC ₅₀ by MLR	Prd. log IC ₅₀ by PLS
A3	77	7.11	6.90	6.49	6.51	B15	83	7.08	7.29	8.22	8.04
A24*	399	6.4	6.30	6.88	6.85	B18	5.6	8.25	8.43	8.37	8.22
A25*	265	6.58	6.93	7.12	7.22	B19	1.6	8.8	8.59	8.50	8.31
A26	132	6.88	7.09	7.10	7.04	B24*	1.8	8.74	8.94	8.71	8.48
A27	369	6.43	6.48	7.18	7.14	B25*	0.59	9.23	9.47	8.50	8.15
A28	917	6.04	6.25	7.08	7.10	B26	0.54	9.27	9.06	8.00	7.87
A29	601	6.22	6.43	7.09	7.10	B27*	0.11	9.96	9.37	8.22	8.01
A31	2300	5.64	6.42	6.90	6.89	B32	3.2	8.49	8.55	8.18	8.07
A32	5400	5.27	6.40	7.14	7.11	B33	0.9	9.05	9.26	8.53	8.18
A33	777	6.11	6.56	6.87	6.77	B34	1.1	8.96	8.75	7.98	7.89
A34	330	6.48	6.62	6.81	6.91	B35	0.18	9.74	9.53	8.26	7.75
A35	120	6.92	6.71	7.06	7.00	C1	31	7.51	7.72	7.27	7.48
A36*	344	6.46	6.19	6.71	6.80	C4	920	6.04	6.25	7.72	7.87
A37	89	7.05	6.91	7.23	7.14	C8	61	7.21	7.50	7.77	8.04
A38*	53	7.28	7.35	7.19	7.05	C9	38	7.42	7.63	7.81	8.10
A39	130	6.89	7.02	7.30	7.21	C10	3.4	8.47	8.26	7.86	8.22
A48	192	6.72	6.57	6.92	6.94	C13	17	7.77	7.56	7.76	8.00
A49	29	7.54	7.17	6.47	6.44	C14	4	8.4	7.84	7.85	8.09
A50*	300	6.52	6.92	6.49	6.47	C15*	12	7.92	8.03	7.90	8.09
A51*	101	7	6.10	7.14	7.13	C16	8.2	8.09	8.30	7.98	8.00
A52	32	7.49	7.28	7.10	7.14	C17	4.8	8.32	8.53	7.93	8.25
A67	210	6.68	6.89	6.61	6.66	C18	6.3	8.2	8.25	8.60	8.19
A68*	150	6.82	7.00	7.19	7.24	C19*	1.7	8.77	9.62	8.57	8.40
A69*	57	7.24	6.75	6.70	6.93	D1	3.8	8.42	8.21	8.76	9.01
A70	67	7.17	6.96	6.63	6.74	D2	40	7.4	7.61	8.03	7.98
A71	168	6.77	6.98	6.70	6.56	D3	8	8.1	8.31	8.16	8.16
A55	27	7.57	7.04	7.13	7.10	D6*	11	7.96	8.76	8.46	8.62
A56	51	7.29	7.36	7.08	7.12	D7*	28	7.55	8.51	9.45	9.02
A58	19	7.72	6.71	6.72	6.68	D8	1	9	9.21	8.04	8.27
A75	254	6.6	6.62	6.80	6.81	D9	13	7.89	8.10	8.54	8.60
A76	223	6.65	6.67	6.86	6.93	D12	0.7	9.15	8.94	8.32	8.32
A77	102	6.99	7.16	6.30	6.22	E4	340	6.47	6.63	6.79	6.82
A78*	256	6.59	6.06	7.05	7.11	E5	3606	5.44	5.65	6.70	6.66
A79	230	6.64	6.75	6.83	6.80	E6	241	6.62	6.83	6.79	6.88
A80	214	6.67	6.80	7.02	7.00	E7	72	7.14	6.93	6.84	6.84
A81	37	7.43	7.17	6.90	6.89	E8*	71	7.15	6.98	6.98	7.04
A82	94	7.03	6.82	6.59	6.62	E9	13	7.89	7.87	7.11	7.42
A83	158	6.8	7.06	6.91	6.90	E11	140	6.85	6.64	5.69	5.79
A84	39	7.41	7.20	6.96	7.04	E15	128	6.89	7.10	7.93	7.58
A87	57	7.24	6.75	6.98	6.82	E10	549	6.26	6.47	5.85	6.00
A97	140	6.85	6.89	6.52	6.51	E12	10	8	8.21	7.93	7.72
A98	2290	5.64	6.80	6.48	6.44	E16	9	8.05	8.26	8.15	8.13
A99	187	6.73	6.94	6.33	6.36	E17	3	8.52	8.31	8.06	7.96
A100	2600	5.59	5.80	6.78	6.80	E18	6	8.22	8.43	8.48	8.53
A101	81	7.09	7.30	6.33	6.34	E19*	7	8.15	8.66	8.27	8.44
B36	917	6.04	6.25	6.84	7.11	E20	15	7.82	7.61	7.69	7.58
B2	5	8.3	8.29	8.39	8.38	E21	0.5	9.3	9.09	8.93	8.77
B10*	142	6.85	7.54	8.11	7.94	E22	0.5	9.3	9.51	8.86	8.85
B11	36	7.44	7.55	8.14	8.03	E23	0.4	9.4	9.19	8.54	8.59

Compounds marked with an asterisk pertain to the test set.

Table 2 Various statistical parameters and figures of merit of the developed models

Para.	SVM		MLR		PLS	
	Training set	Testing set	Training set	Testing set	Training set	Testing set
N	79	19	79	19	79	19
RMSE	0.316	0.276	0.685	0.544	0.703	0.547
R ²	0.935	0.728	0.553	0.464	0.529	0.461
Q ²	0.905	—	0.404	—	0.441	—

the dependent variable. The regression coefficients represent the independent contribution of each molecular descriptor. The MLR model is defined by the following matrix representation:

$$y = bX + e$$

When X is of full rank, the least squares solution is:

$$\hat{b} = (X^T \cdot X)^{-1} X^T y$$

Here, $\hat{b}$ represents the estimator for the regression coefficients in $\hat{b}$.

The predicted activities of pIC₅₀ of the training and test data and the relative error (RE) of the model prediction are listed in Table 2.

Partial Least Squares (PLS) method

Parallel to the SMLR, another linear model was established by applying the partial least squares (PLS) tool to the original data matrix. PLS is a regression method for establishing a prediction model between two variable matrices: the X matrix as the predictor variables and the Y matrix as the activity variables. It is assumed that the X matrix ($I \times J$) contains the descriptors that can be used to predict the active matrix, which is y ($I \times M$). Here, the dependent variable is represented by an ($I \times 1$) column vector. PLS decomposes these matrices into a two matrix product plus residuals.

$$X = TP^T + E = \sum t_j p'_j + E$$

$$y = UQ^T + F = \sum u_j q'_j + F$$

Here, T and U represent the scoring matrices of X and y ; P and Q are the load matrices of X and y ; E and F are the matrix residuals of the model with f latent variables. The above equations are solved in a way that the covariance between T and U is maximized. The two matrices are related by the following internal relationships:

$$U = TB + H$$

Here, B is the diagonal matrix and H is the residual matrix. This allows PLS to be characterized as a predictive model.

The matrix y can be calculated using U as follows:

$$y = TBQ^T + F$$

The activities of the new compounds can be estimated from the new score T^* , which in the above equation leads to the following equation:

$$Y_{\text{pred}} = T^* B Q^T$$

In the partial least squares method, the number of latent variables (LVs) is crucial. In the model development, using an insufficient amount of LVs will result in an under-fitting model, while the use of excessive LVs tends to produce over-fitting models. Therefore, RMSE_{CV} and RMSE_C are used to select the optimal number of latent variables.

The predicted activities of pIC₅₀ of the training and test data and relative error (RE) of prediction using the model can be found listed in Table 2.

Support Vector Machine (SVM) method

QSAR analysis typically generates more descriptors than those required for the construction of the prediction model (Table S1, ESI†). In an effort to select the most relevant descriptors, the SVM feature selection tool²⁸ was applied. SVMs can be used to carry out the general regression and classification (of η and ε -type), as well as density-estimation. Here, the basic idea of SVR is to map the data x into a higher-dimensional feature space F via nonlinear mapping. Then, linear regression can be applied in this space. The major advantage of SVR over other methods is that it adopts a structure risk minimization (SRM) principle. In the present study, SVR was implemented by the SVM function in the e1071 R package,³¹ which supports several kernel functions (radial basis or Gaussian, polynomial, sigmoid, and linear functions). In the present study, we employed radial basis function (RBF) kernels in the modeling. During the QSAR modeling, the leave-one-out cross validation was employed to find a promising QSAR model. Finally, the cross-validation correlation coefficient, a measure of the model generalization capability for all candidate models, was obtained.

The predicted activities of pIC₅₀ of the training and test data and relative error (RE) of the prediction model are listed in Table 2.

Evaluation of model performance

In machine learning, the assessment of prediction quality of classification models is typically performed on the basis of the error rate and the overall accuracy. In the present study, we adopt q_{ed}^2 , a ten-fold cross validation and external set validation to test the performance of the models. In addition, ten-fold cross validation of the training sets was performed to evaluate the robustness of the models.

Molecular dynamics (MD) simulations

The docked structures of an inhibitor complexed with B-Raf^{V600E} were used as the initial structures for MD calculations. A CHARMM force field was applied to the complex and the resulting system was then immersed into an "Explicit Periodic Boundary" water box with a sodium cation used for complex neutralization. Afterwards, the solvated system was subjected to double-fold minimization (10 000 cycles of steepest descent minimization and 100 000 cycles of conjugate gradient minimization). The system was gradually heated from 50 K to 300 K over a period of 100 ps and subsequently equilibrated for 500 ps. Starting from the last frame of the equilibration, production simulation was performed for 500 ps using the NPT ensemble under a constant temperature of 300 K and pressure of 1 atm. Other parameters of the MD simulation were maintained at the default Discovery Studio configuration.

Interaction energy analysis

In order to explore the inhibitor-protein interaction mode and highlight the key residues responsible for the binding affinity,

interaction energies for each inhibitor were further decomposed into individual residue contributions using the "Calculate Interaction Energy" protocol. The stable coordinates extracted in the last 500 ps MD simulation were employed here for the interaction energy decomposition analysis.

Results and discussion

Development and advantage of the MD-SVR model

The structures of the 98 molecules were generated and a large number of descriptors (X-segment columns, various score function evaluations) (Fig. 1) were obtained by docking each molecule with B-Raf using the corresponding molecular structure. In order to determine the relationship between the structural binding pattern as a variable (ligand fit scores and combination parameters) and the biological activities as independent variables, the logarithm of the reciprocal biological activity of the 98 molecules ($-\log IC_{50}$) was used. In this study, three regression methods were applied sharing the same descriptors, including two linear regression methods, MLR and PLS, and a nonlinear regression method, SVM. The corresponding results for each regression method are shown listed in Table 2. Besides, the predictive activities of the pIC_{50} for the training and test data and the relative error (RE) of the model prediction are listed in Table 1. The low RE confirms the high predictability of the model. The pIC_{50} values estimated by various modeling methods *versus* experimental values for the training and test sets are shown in Fig. 2. Furthermore, the plots of residuals *versus* experimental factors for each method are shown in Fig. 3.

In general, the purpose of multivariate linear regression is to quantify the relationship between several independent or predictor variables and the dependent variable. The independent or predictor variables can be a particular plurality of physicochemical descriptors of a molecule, their major components, or potential variables. Compared to the previously reported two regression models, the support vector machine classification (SVC) method presents a one-to-one method for solving multi-class problems.³² The main advantage of the support vector machine³³ is that it uses a Structural Risk Minimization (SRM) principle, which has been shown to be superior to the traditional empirical risk minimization (ERM)

principles used in traditional neural networks. In our MD-SVR modeling, the RBF kernel was used as the kernel function.³⁴ Therefore, the capacity parameters C and ε of ε -insensitive loss functions as well as the corresponding parameters γ of RBF kernel need to be optimized. Noteworthy in this context is the finding that the best cross-validation result (error: 0.1972, Fig. 4) was obtained with optimal values of C , γ and $\varepsilon = 5$, 0.2 and 0.05, respectively. From this data we were able to determine that the result of the SVM regression was more accurate than those of the MLR and PLS prediction. For the training set, $R^2 = 0.9348$ and $RMSE = 0.3165$, and for the cross validation, $Q^2 = 0.9049$. When the model was applied to the test set, values of $R_T^2 = 0.7387$ and $RMSE_T = 0.2758$ were obtained.

Prediction ability of the MD-SVR model in the screening of novel B-Raf inhibitors

In order to validate the practical application ability of the established MD-SVR models in the development of novel B-Raf inhibitors, a natural product database screening was carried out based on the obtained MD-SVR model. Two natural products, quercetin and myricetin, were screened and exhibited promising predicted activities (Table 3). Furthermore, the enzyme inhibitory activities of the compounds against B-Raf^{V600E} were biologically evaluated, with commercial GW5074 as a positive control. Although the predicted activity of quercetin (pred. $IC_{50} = 7.41$ nM) was superior to that of myricetin (pred. $IC_{50} = 66.07$ nM), the activity test showed that the inhibitory potency of myricetin (exp. $IC_{50} = 1.56$ μ M) to B-Raf was 5-fold higher than that of quercetin (exp. $IC_{50} = 7.59$ μ M), with a significant difference compared to the positive control GW5074. Indeed, preliminary screening has been demonstrated to be sufficient for further structural transformation as a lead molecule.

In order to provide a structural explanation for the above experimental results, molecular docking analysis was performed to further explore the binding mode of the newly screened derivatives (Fig. 5). Furthermore, the lipophilic potential (LP), hydrogen bonding site (HB) and electrostatic potential (EP) could be determined. To further investigate the interaction between quercetin and myricetin with the receptor B-Raf, docking

Fig. 2 pIC_{50} estimated by various modeling methods *versus* experimental values for the training and test sets: (a) SVM (b) MLR (c) PLS.

Fig. 3 Plots of residuals versus experimental factors: (a) SVM (b) MLR (c) PLS.

studies were performed. The docking results exhibited that both flavone groups in quercetin and myricetin may help the compounds occupy the hydrophobic pocket of B-Raf. Furthermore, the flavone unit is able to form hydrogen bonds with GLN530. We also found that their 3',4'-hydroxyl groups are able to form hydrogen bonds with LYS483, strengthening the overall binding affinity of quercetin and myricetin to B-Raf. In addition, we determined that myricetin could bind to B-Raf more tightly than quercetin, a finding that was consistent with the corresponding inhibitory potencies.

MD simulations and interaction mechanism analysis

In order to investigate the interaction mode between the preferable compound quercetin (or myricetin) and the B-Raf^{V600E} protein, molecular dynamics (MD) simulations and interaction decomposition analysis were carried out to explore the quantitative energy contributions per residue to the binding affinity of the promising inhibitor (myricetin). The RMSD value of B-Raf reaches equilibration and swings around at an average value after a 400 ps simulation time. The RMSD value of the protein backbone was calculated from a 100 to 400 ps trajectory, where the data points fluctuated by 1.3499 ± 0.09 nm. After 400 ps production simulation, the distance between the inhibitor (myricetin) and key amino acid Lys591 of B-Raf tends to converge, indicating that the system reached equilibrium conditions (Fig. 6). Based on the stable conformation obtained from the MD simulation, interactions involved in the protein/ligand complex were analysed. As shown in Fig. 6(d), the carbonyl and hydroxyl groups on the catechin of myricetin formed hydrogen bonds with the NH and carbonyl groups of B-Raf hinge residue LYS591 located in the hydrophobic pocket of B-Raf. In accordance with the interaction energy decomposition, the contribution of LYS591 for myricetin is -32.09 kcal mol⁻¹ (Fig. 6(b)). Therefore, according to the binding pattern and chemical structure of myricetin, we speculate that the aromatic moiety, linked by a catechin scaffold containing hydrogen-donors and acceptors, may be an important pharmacophore feature of B-Raf inhibitors.

Conclusions

In this manuscript, we describe the development and validation of molecular descriptor- and docking-based models for the computational identification of a series of B-Raf^{V600E}

Fig. 4 (a) The 3D plot of cross-validation error for the SVM when choosing the optimal parameters γ , ϵ and C . (b) The 3D plot of the cross-validation error for the SVM.

Table 3 The predictive and tested IC₅₀s against B-Raf of the screened natural products quercetin and myricetin

Compound	Pred.-pIC ₅₀	Pred. IC ₅₀ (nM)	Experimental IC ₅₀ (μM)
Quercetin	8.13	7.41	7.589
Myricetin	7.18	66.07	1.562
GW5074	—	—	0.0015

Fig. 5 The docking conformation of quercetin (a) and myricetin (b) in the active site of the B-Raf kinase domain where the protein is shown in its secondary structure and quercetin and myricetin are represented as a ball and stick model.

inhibitors. Three regression methods were applied in order to establish reliable and structure-based QSAR models. The optimal method molecular docking based SVR was used as a reasonable virtual screening method to obtain novel structures of B-Raf^{V600E} inhibitors. The screened compounds quercetin and myricetin, obtained from a natural product database (NPD from the ZINC database), were screened for their inhibitory activities. Both compounds exhibited promising B-Raf^{V600E} inhibitory ($IC_{50}^{\text{Quercetin}} = 7.59 \mu\text{M}$ and $IC_{50}^{\text{Myricetin}} = 1.56 \mu\text{M}$) activities, demonstrating the high reliability and good applicability of the established molecular docking based SVR model in the development of B-Raf^{V600E} inhibitors. Lastly, the binding mode of myricetin with the protein was studied *via* docking as well as MD simulations, and the compound myricetin was determined to be a good lead candidate for further structure-activity studies in the search for the most potent inhibitors.

Experimental

In vitro biological assays

The inhibitory activity of the compounds against the B-Raf^{V600E} enzyme was evaluated according to a published procedure.²⁷ Specifically, a 9.9 μL reaction in assay buffer (50 mM HEPES, pH = 7.5, 10 mM MgCl_2 , 1 mM EGTA, 0.01% BRIJ-35) containing 0.6 nM B-Raf^{V600E}, 0.2 μM fluorescein-MAP2K1, and 1.5 μM ATP was carried out using a 384-well plate. The reaction mixture was then incubated at room temperature for 1 hour and subsequently quenched with 10 μL of a detection solution containing 2 nM antibody and 10 mM

Fig. 6 MD simulations and interaction mechanism analysis for the most potent inhibitor myricetin during a 400 ps simulation time: (a) RMSD of the B-Raf backbone, (b) plot of the binding pattern of myricetin, (c) plot of the distances of the H-bond between Lys591 and myricetin and (d) the binding interaction of myricetin with B-Raf.

EDTA. After incubation for 30 min, the reaction mixture was analyzed using a Caliper Lab Chip 3000 (Caliper Life Sciences, Hopkinton, MA, USA) by electrophoretic separation of the fluorescent substrate and phosphorylated product. All compounds were first evaluated at a concentration of 10 μ M and further screened for seven concentrations prepared from 3-fold serial dilutions with a preliminary activity of more than 40%. The IC₅₀ values were obtained using the analytical software Prism 5.0 (GraphPad Software Pte Ltd.).

Acknowledgements

This study was financially supported by the National Natural Science Foundation of China (81673294, 81502914 and 81502926) and the Science and Technology Planning Project of Zhejiang Province (2016C33067).

Notes and references

- G. Bollag, P. Hirth, J. Tsai, J. Z. Zhang, P. N. Ibrahim, H. N. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. J. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, K. D'Andrea, A. Koehler, M. Stumm, P. S. Lin, R. J. Lee, J. Grippo, I. Puzanov, K. B. Kim, A. Ribas, G. A. McArthur, J. A. Sosman, P. B. Chapman, K. T. Flaherty, X. W. Xu, K. L. Nathanson and K. Nolop, *Nature*, 2010, **467**, 596–599.
- X. Q. Kong, J. Qin, Z. Li, A. Vultur, L. J. Tong, E. G. Feng, G. Rajan, S. E. Liu, J. Y. Lu, Z. J. Liang, M. Y. Zheng, W. L. Zhu, H. L. Jiang, M. Herlyn, H. Liu, R. Marmorstein and C. Luo, *Org. Biomol. Chem.*, 2012, **10**, 7402–7417.
- E. R. Cantwell-Dorris, J. J. O'Leary and O. M. Sheils, *Mol. Cancer Ther.*, 2011, **10**, 385–394.
- M. A. Rahman, A. Salajegheh, R. A. Smith and A. K. Y. Lam, *Crit. Rev. Oncol. Hematol.*, 2014, **90**, 220–232.
- M. M. Vasbinder, B. Aquila, M. Augustin, H. Chen, T. Cheung, D. Cook, L. Drew, B. P. Fauber, S. Glossop, M. Grondine, E. Hennessy, J. Johannes, S. Lee, P. Lyne, M. Mortl, C. Omer, S. Palakurthi, T. Pontz, J. Read, L. Sha, M. Shen, S. Steinbacher, H. Wang, A. Wu and M. Ye, *J. Med. Chem.*, 2013, **56**, 1996–2015.
- P. Xie, C. Streu, J. Qin, H. Bregman, N. Pagano, E. Meggers and R. Marmorstein, *Biochemistry*, 2009, **48**, 5187–5198.
- A. Zambon, I. Niculescu-Duvaz, D. Niculescu-Duvaz, R. Marais and C. J. Springer, *Bioorg. Med. Chem. Lett.*, 2012, **22**, 789–792.
- B. Gencler and M. Gonul, *Dermat. Res. Pract.*, 2016, **2016**, 1–6.
- Z. Wang, H. Y. Sun, X. J. Yao, D. Li, L. Xu, Y. Y. Li, S. Tian and T. J. Hou, *Phys. Chem. Chem. Phys.*, 2016, **18**, 12964–12975.
- R. Arora, M. Di Michele, E. Stes, E. Vandermarliere, L. Martens, K. Gevaert, E. Van Heerde, J. T. M. Linders, D. Brehmer, E. Jacoby and P. Bonnet, *J. Med. Chem.*, 2015, **58**, 1818–1831.
- S. X. Dai, W. X. Li, F. F. Han, Y. C. Guo, J. J. Zheng, J. Q. Liu, Q. Wang, Y. D. Gao, G. H. Li and J. F. Huang, *Sci. Rep.*, 2016, **6**.
- S. K. Venkatesan, P. Saudagar, A. K. Shukla and V. K. Dubey, *Interdiscip. Sci.: Comput. Life Sci.*, 2011, **3**, 217–231.
- C. Bobach, S. Tennstedt, K. Palberg, A. Denkert, W. Brandt, A. de Meijere, B. Seliger and L. A. Wessjohann, *Eur. J. Med. Chem.*, 2015, **90**, 267–279.
- E. Vrontaki, G. Melagraki, T. Mavromoustakos and A. Afantitis, *Methods*, 2015, **71**, 4–13.
- C. Cortes and V. Vapnik, *Mach. Learn.*, 1995, **20**, 273–297.
- T. T. Yao, J. L. Cheng, B. R. Xu, M. Z. Zhang, Y. Z. Hu, J. H. Zhao and X. W. Dong, *RSC Adv.*, 2015, **5**, 49195–49203.
- S. Zhang, A. Golbraikh, S. Oloff, H. Kohn and A. Tropsha, *J. Chem. Inf. Model.*, 2006, **46**, 1984–1995.
- X. W. Dong, C. Y. Jiang, H. Y. Hu, J. Y. Yan, J. Chen and Y. Z. Hu, *Eur. J. Med. Chem.*, 2009, **44**, 4090–4097.
- X. W. Dong, Y. M. Zhao, X. Q. Huang, K. N. Lin, J. Z. Chen, E. Q. Wei, T. Liu and Y. Z. Hu, *Eur. J. Med. Chem.*, 2013, **62**, 754–763.
- W. H. Zhan, D. Q. Li, J. X. Che, L. R. Zhang, B. Yang, Y. Z. Hu, T. Liu and X. W. Dong, *Eur. J. Med. Chem.*, 2014, **75**, 11–20.
- X. W. Dong, X. L. Zhou, H. Jing, J. Z. Chen, T. Liu, B. Yang, Q. J. He and Y. Z. Hu, *Eur. J. Med. Chem.*, 2011, **46**, 5949–5958.
- T. T. Yao, J. F. Xie, X. G. Liu, J. L. Cheng, C. Y. Zhu, J. H. Zhao and X. W. Dong, *RSC Adv.*, 2017, **7**, 10353–10360.
- A. Golbraikh and A. Tropsha, *J. Mol. Graphics Modell.*, 2002, **20**, 269–276.
- S. Mathieu, S. N. Gradl, L. Ren, Z. Wen, I. Aliagas, J. Gunzner-Toste, W. Lee, R. Pulk, G. Zhao, B. Alicke, J. W. Boggs, A. J. Buckmelter, E. F. Choo, V. Dinkel, S. L. Gloor, S. E. Gould, J. D. Hansen, G. Hastings, G. Hatzivassiliou, E. R. Laird, D. Moreno, Y. Ran, W. C. Voegtli, S. Wenglowsky, J. Grina and J. Rudolph, *J. Med. Chem.*, 2012, **55**, 2869–2881.
- L. Ren, S. Wenglowsky, G. Miknis, B. Rast, A. J. Buckmelter, R. J. Ely, S. Schlachter, E. R. Laird, N. Randolph, M. Callejo, M. Martinson, S. Galbraith, B. J. Brandhuber, G. Vigers, T. Morales, W. C. Voegtli and J. Lyssikatos, *Bioorg. Med. Chem. Lett.*, 2011, **21**, 1243–1247.
- S. Wenglowsky, L. Ren, K. A. Ahrendt, E. R. Laird, I. Aliagas, B. Alicke, A. J. Buckmelter, E. F. Choo, V. Dinkel, B. Feng, S. L. Gloor, S. E. Gould, S. Gross, J. Gunzner-Toste, J. D. Hansen, G. Hatzivassiliou, B. Liu, K. Malesky, S. Mathieu, B. Newhouse, N. J. Raddatz, Y. Ran, S. Rana, N. Randolph, T. Risom, J. Rudolph, S. Savage, L. T. Selby, M. Shrag, K. Song, H. L. Sturgis, W. C. Voegtli, Z. Wen, B. S. Willis, R. D. Woessner, W. I. Wu, W. B. Young and J. Grina, *ACS Med. Chem. Lett.*, 2011, **2**, 342–347.
- T. R. Rheault, J. C. Stellwagen, G. M. Adjabeng, K. R. Hornberger, K. G. Petrov, A. G. Waterson, S. H. Dickerson, R. A. Mook, Jr., S. G. Laquerre, A. J. King, O. W. Rossanese, M. R. Arnone, K. N. Smitheman, L. S. Kane-Carson, C. Han, G. S. Moorthy, K. G. Moss and D. E. Uehling, *ACS Med. Chem. Lett.*, 2013, **4**, 358–362.

- 28 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- 29 S. Mathieu, S. N. Gradl, L. Ren, Z. Y. Wen, I. Aliagas, J. Gunzner-Toste, W. Lee, R. Pulk, G. L. Zhao, B. Aliche, J. W. Boggs, A. J. Buckmelter, E. F. Choo, V. Dinkel, S. L. Gloor, S. E. Gould, J. D. Hansen, G. Hastings, G. Hatzivassiliou, E. R. Laird, D. Moreno, Y. Q. Ran, W. C. Voegtli, S. Wenglow sky, J. Grina and J. Rudolph, *J. Med. Chem.*, 2012, **55**, 2869–2881.
- 30 M. Montes, M. A. Miteva and B. O. Villoutreix, *Proteins*, 2007, **68**, 712–725.
- 31 Y. Li, Y. Kong, M. D. Zhang, A. X. Yan and Z. M. Liu, *Mol. Inf.*, 2016, **35**, 116–124.
- 32 C. C. Chang and C. J. Lin, *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, 1–27.
- 33 V. Vapnik and A. Lerner, *Autom. Remote Control*, 1963, **24**, 774–780.
- 34 Q. Zang, D. M. Rotroff and R. S. Judson, *J. Chem. Inf. Model.*, 2013, **53**, 3244–3261.