

Cite this: *Soft Matter*, 2011, **7**, 4001

www.rsc.org/softmatter

PAPER

# Stroke prevention by traditional Chinese medicine? A genetic algorithm, support vector machine and molecular dynamics approach†

Kuan-Chung Chen<sup>a</sup> and Calvin Yu-Chian Chen<sup>\*abcd</sup>

Received 27th December 2010, Accepted 20th January 2011

DOI: 10.1039/c0sm01548b

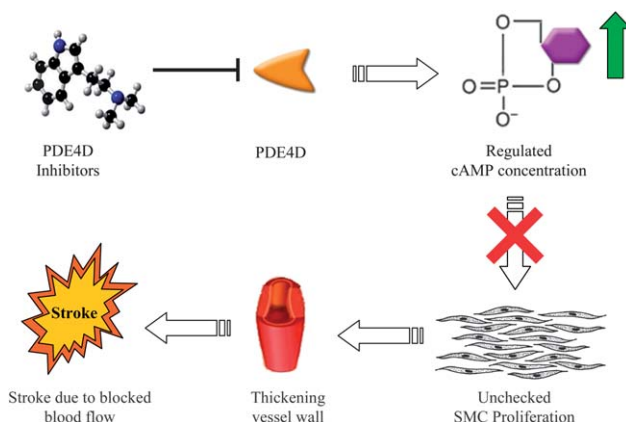
Phosphodiesterase 4D (PDE4D) has been identified as a promising target which associates with stroke, which is one of the top 3 leading causes of death and main leading cause of adult disability in US. In this study, we applied virtual screening on the world's largest traditional Chinese medicine (TCM) database (<http://tcm.cmu.edu.tw>;<sup>1</sup> C. Y. C. Chen, *PLoS One*, 2011, **6**, e15939) for natural compounds that inhibit PDE4D functions. Molecular docking and dynamics simulation were employed to investigate the protein–ligand interactions of compounds with top two dock scores. During the simulation, the divalent metal cations in PDE4D formed stable hydrogen bonds and electrostatic interactions between ligand and binding site residues. Furthermore, the two top TCM candidates, 2-*O*-caffeoyl tartaric acid and mumefural, formed additional steady hydrogen bonds with binding site residue and active site residue respectively. The additional hydrogen bonds further stabilize protein–ligand interaction at the PDE4D binding site. To predict the bioactivity of the top TCM candidates, we built two prediction models using multiple linear regression (MLR) and support vector machine (SVM). The predicted pIC<sub>50</sub> values suggest that 2-*O*-caffeoyl tartaric acid and mumefural are potential PDE4D inhibitors.

## Introduction

Stroke and cerebrovascular diseases are the top 3 leading causes of death in U.S. in last several years.<sup>2</sup> Moreover, stroke is the main leading cause of adult disability.<sup>3</sup> Stroke is a multifactorial disease that can have both genetic and environmental contributions. In the past few years, researchers have attempted to explore the polygenic factors behind stroke by using genome-wide association study; however, the immense amount of information stored in human genome has presented arduous logistical and technical challenges.

Recently, the Icelandic Decode group has identified a novel stroke-related gene, phosphodiesterase 4D (PDE4D), by using genome-wide association screen.<sup>4</sup> Since the study from Gretarsdottir, many replication studies have been conducted using different population samples, including Staton,<sup>5</sup> Sun,<sup>6,7</sup> Nakayama,<sup>8</sup> van Rijn,<sup>9</sup> Saleheen,<sup>10</sup> Nilsson-Ardnor,<sup>11</sup> Bevan,<sup>12</sup> and Brophy,<sup>13</sup> in an attempt to examine the relationship between PDE4D and stroke.

PDE4D is a cyclic nucleotide phosphodiesterase which selectively hydrolyzes cyclic AMP (cAMP) to control the concentration of second messenger cAMP. The concentration of cAMP influences various cellular metabolisms. Decreasing in cAMP concentration could lead to increasing proliferation and migration of smooth muscle cell that subsequently could result in development of atherosclerosis.<sup>4</sup> Therefore, a PDE4D-targeting drug which is designed to obstruct PDE4D activation can reduce the risk of stroke, like the design shown in Fig. 1. This study



**Fig. 1** Schematic diagram representing the proposed drug design strategy. By inhibiting PDE4D activity, the cAMP concentration will rise and hence prevent over proliferation of smooth muscle cell (SMC). Thickening blood vessel wall that blocks blood flow and causes stroke can hence be prevented by regulating SMC growth.

<sup>a</sup>Laboratory of Computational and Systems Biology, School of Chinese Medicine, China Medical University, Taichung, 40402, Taiwan. E-mail: ycc@mail.cmu.edu.tw; ycc929@MIT.EDU; Tel: +1-617-353-7123

<sup>b</sup>Department of Bioinformatics, Asia University, Taichung, 41354, Taiwan

<sup>c</sup>Department of Systems Biology, Harvard Medical School, Boston, MA, 02115, USA

<sup>d</sup>Department of Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c0sm01548b

focused on identifying potential lead compounds that could act as PDE4D inhibitor.

Large structural diversity found in Chinese herbal ingredients and natural products poses a very valuable resource for finding novel lead compounds. Various natural compounds have been shown to be useful in anti-cancer,<sup>14,15</sup> anti-radical, anti-inflammatory and anti-platelet therapies.<sup>16</sup> We, therefore, employed a traditional Chinese medicine (TCM) database (<http://tcm.cmu.edu.tw>)<sup>1</sup> and natural compounds for virtual screening by molecular docking. Molecular dynamics simulations were performed after docking to study receptor-ligand interactions and to determine stabilized conformations for the obtained protein-ligand complexes. Currently, bioactivity prediction using machine learning and building QSAR model is becoming a routine process in drug design; some novel examples include using support vector machine (SVM) for identifying bioactive natural products<sup>17</sup> and building quantitative structure–activity relationship (QSAR) models for predicting inhibitory activity of small molecules.<sup>18–27</sup> To predict the bioactivities of TCM candidates obtained from screening, we also constructed QSAR models using both multiple linear regression approach and SVM approach. The prediction resulted from multiple linear regression (MLR) and SVM can further support the efficacy of the TCM candidates.

## Materials and methods

### Data set

The molecular simulations were performed by using Accelrys Discovery Studio 2.5 (DS 2.5). A total of 20,000 TCM compounds were downloaded from TCM database (<http://tcm.cmu.edu.tw>).<sup>1</sup> All the downloaded TCM compounds had been minimized in MM2 force field using Chem 3D Ultra 11.0. The crystal structure of PDE4D (PDB ID: 3G4G)<sup>28</sup> was obtained from RCSB Protein Data Bank. Known PDE4D inhibitor, RS-253344, which co-crystallized in the PDE4D cAMP binding pocket, was used as a control.

### Molecular docking

The docking simulation was performed in the force field of CHARMM<sup>29</sup> using LigandFit module<sup>30</sup> in the Receptor-Ligand Interactions package of DS 2.5. The compounds from TCM Database were docked into the binding site of PDE4D which was defined by the volume of RS-25344 present in the initial crystal structure. Additionally, the inhibitor RS-25344 and cAMP, the normal substrate for PDE4D, were used as controls. The receptor was fixed in the docking protocol, but the ligands that complement the binding site were initially flexible but were later fixed for rigid-body energy minimization after docking. All chemical compounds and their possible poses were evaluated by scoring functions. In addition, top ranking compounds were visually inspected for binding poses, binding location and interacting residues. Our selection criteria specified the potential candidates to have similar binding characteristics and locations with the control molecules.

Dock Score that predicts ligand binding affinities by calculating ligand internal energy and receptor-ligand interaction energy was used as the primary scoring function.<sup>30</sup> Other scoring

functions, PLP, PLP2 and PMF, were also calculated for additional references but were not used in selecting ligand and their poses. The PLP and PLP2 stand for piecewise linear potentials which predict protein-ligand binding affinities by calculating hydrogen bond interaction.<sup>31</sup> These two versions of pairwise linear potentials differ in that PLP2 introduces atomic radii and angular components. The PMF (potential of mean force) predicts ligand binding affinities by summing all interatomic pairwise interactions of the receptor-ligand complex.<sup>32</sup>

### Molecular dynamics simulation

The molecular dynamics (MD) simulation was performed using Standard Dynamics Cascade module and Dynamics (Production) module in the Simulation package of DS 2.5 with CHARMM.<sup>29</sup> The complexes were created a 7 Å solvation shell around the protein. Each system was also neutralized by additional sodium cations. Two cycles of minimization have been performed, each using different system condition. First, the minimization was performed with the atoms of protein restrained. Afterwards, the minimization was performed without restraint. The minimization was performed using both Steepest Descent<sup>33</sup> and Conjugate Gradient<sup>34</sup> with maximum up to 6000 cycles. The SHAKE algorithm was applied to fix all bonds involving hydrogen atoms throughout the MD simulation, and the long-range electrostatics were treated with PME method. The time step was set to 0.001 ps for all MD stages. After minimization, the protein-ligand complex was gradually heated to target temperature of 310 K from 51 K over an interval of 50 ps. After heating procedure and a 200 ps equilibration phase, the production stage was performed for 20 ns using NVT canonical ensemble. The temperature coupling decay time for the Berendsen thermal coupling method was 0.4 ps. Post processing of the trajectory was done using Analyze Trajectory module at Simulation package of DS 2.5.

### Support vector machine and multiple linear regression

To predict the bioactivity of the TCM compounds, we built activity prediction models using both SVM approach and MLR approaches. These two approaches differ in that SVM utilizes a non-linear feature mapping technique whereas MLR uses linear evaluation. The SVM model in the present study was built using LibSVM.<sup>35</sup> Multiple linear regressions were performed using MATLAB.

The 62 compounds used in building the predictive model were compiled from three different sources<sup>36–38</sup> and were divided randomly into training set and test. These compounds were drawn using ChemBioOffice 2008 and prepared in Discovery Studio 2.5 using the Prepare Ligand module to modify the compound ionization state to physiological ionization setting. The activity (IC<sub>50</sub>) data used in QSAR were converted to pIC<sub>50</sub>.

The molecular descriptors were calculated using Calculate Molecular Properties in Discovery Studio 2.5, which could calculate 552 different descriptors. To select molecular descriptors that best represent the training set molecules, the genetic function approximation (GFA)<sup>39</sup> was used. The GFA algorithm is based on genetic algorithm, which simulates a population of strings based on the concept of natural evolution. In genetic

algorithm, selection, crossover and mutation events are iteratively simulated for a specified number of generations. By applying genetic algorithm, the GFA searches all possible QSAR models and then uses square correlation coefficient ( $R^2$ ) to estimate the fitness of individual model. The cross validation test is applied during building model.

Support vector machine (SVM) is a novel supervised machine-learning method, widely used in bioinformatics.<sup>40–43</sup> SVM minimizes an upper bound of the generalization error on Vapnik-Chernoverkis dimension. This convention of minimization is superior in comparison to the traditional empirical risk minimization (ERM), which only minimizes the training error. In addition to binary classification, SVM can be used in regression of continuous data.<sup>44</sup> This is called support vector regression (SVR). In our cases, the activities ( $pIC_{50}$ ) of inhibitors are continuous. Hence, in order to build a model that can accurately predict inhibitor activity, we used SVR instead of SVM.

The basic idea in SVR is to find the hyperplane that best predicts data distribution by mapping the input data set into a high-dimensional feature space using Kernel function. For a given training data  $(x_1, y_1), \dots, (x_l, y_l) \in \mathbb{R}^d \times \mathbb{R}$  where  $y_i$  denotes the output values of the input vectors  $y_i$  and the data can be expressed in the form of  $f(x) = \langle w, x \rangle + b$ ,  $w \in \mathbb{R}^d$ ,  $b \in \mathbb{R}$ , the SVR finds  $w$  by introducing  $\varepsilon$ -insensitive loss function. This SVR formulation is

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$\text{s.t.} \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \quad (1)$$

This optimization problem can be further transformed to include nonlinear functions by introducing Lagrange multipliers and kernel  $k(x, x') = \langle \phi(x), \phi(x') \rangle$ , which gives

$$\max. \begin{cases} -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) k(x_i, x_j) \\ -\varepsilon \sum_{i=0}^l (\alpha_i - \alpha_i^*) + \sum_{j=0}^l y_j (\alpha_j - \alpha_j^*) \end{cases} \quad (2)$$

$$\text{s.t.} \sum_{i=0}^l (\alpha_i - \alpha_i^*) = 0 \text{ and } \alpha_i - \alpha_i^* \in [0, C]$$

The prediction for new input, in terms of Kernel function, is given by

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) k(x_i, x) + b \text{ s.t. } 0 \leq \alpha_i^* \leq C, 0 \leq \alpha_i \leq C \quad (3)$$

The performance of SVR is significantly dependent on the setting of parameters:  $C$ ,  $\varepsilon$ , the kernel type (mapping function) and the corresponding kernel parameters. In LibSVM, the available kernel functions include linear, polynomial, radial basis function, sigmoid function and precomputed kernel. We constructed the SVR model based on radial basis function, and for our case, the key parameters are  $C$  cost, epsilon and gamma. The  $C$  cost sets the parameter  $C$  in SVR. The epsilon parameter sets  $\varepsilon$  in  $\varepsilon$ -insensitive loss function. The gamma parameter sets the  $\gamma$

value in kernel function. To find the optima values for  $C$ , epsilon, and gamma, we used the gridregression.py program included in the libsvm-2.91 package.

## Results and discussion

### Molecular docking

After virtual screening, the scoring function results of the top 10 compounds and the controls are listed in Table 1. The structures of the top two candidates (2-*O*-caffeoyl tartaric acid and mumeufural) and the two controls (cAMP and RS-25344) are shown in Fig. 2. Scoring functions, Dock Score,<sup>30</sup> was used to rank the docking results. 2-*O*-Caffeoyl tartaric acid and mumeufural both scored higher Dock Score than cAMP and RS-25344 and were, therefore, selected for further analysis.

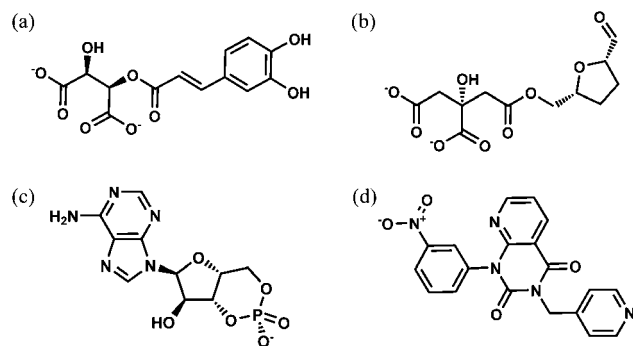
Two-dimensional representation of TCM ligand docking conformations is shown in Fig. 3. In addition, the docking diagrams showed two divalent metal cations,  $Zn^{2+}$  and  $Mg^{2+}$ , which have been suggested as cofactors in cAMP hydrolysis. As illustrated in Fig. 3, both  $Zn^{2+}$  and  $Mg^{2+}$  ions acted as bridges for coordinating ligand interaction to Asp367 and Asp484. All TCM ligands and controls had electrostatic interactions to these two divalent metal cations suggesting the importance of this interaction in ligand binding in PDE4D and the dependency of metal ions in ligand inhibitory activity.

As shown in docking diagrams (Fig. 3), the TCM ligand, 2-*O*-caffeoyl tartaric acid, and two controls had pi-pi interaction with

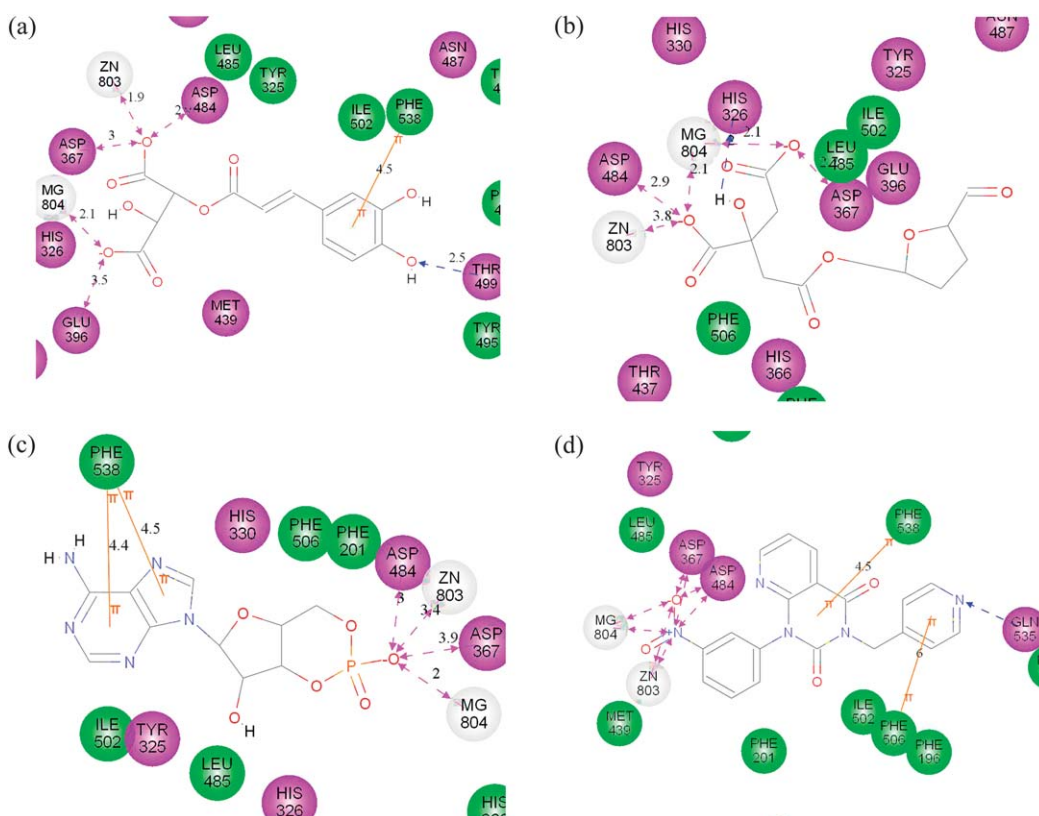
**Table 1** Docking results of top TCM compounds, RS-25344, and cAMP

Name	Dock Score	PLP1	PLP2	PMF
2- <i>O</i> -Caffeoyl tartaric acid	1,013	39.69	34.42	170
Mumeufural	1,009	-12.47	13.82	156
Cyclic AMP <sup>a</sup>	974	32.35	25.43	133
2- <i>O</i> -Feruloyl tartaric acid	948	22.77	39.86	187
Kainic acid	837	-4.26	-3.87	123
Gallic acid	698	28.72	33.29	98
3,5-Dihydroxycinnamic acid	686	16.93	23.24	102
Caffeic acid	686	16.93	23.24	102
2,3-Dihydroxycinnamic acid	676	3.09	8.13	110
Cinnamic acid	672	14.65	18.94	106
Ferulic acid	672	24.28	29.42	110
RS-25344 <sup>a</sup>	250	97.29	73.57	164

<sup>a</sup> Control.



**Fig. 2** The scaffolds of (a) 2-*O*-caffeoyl tartaric acid, (b) mumeufural, (c) cAMP, (d) RS-25344.

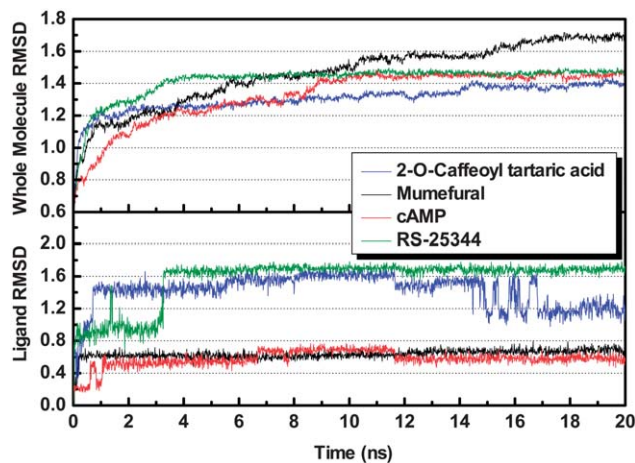


**Fig. 3** Docking poses of (a) 2-*O*-caffeoyl tartaric acid, (b) mumefural, (c) cAMP, (d) RS-25344 in PDE4D cAMP binding site. Electrostatic interaction is shown as dashed line, and pi-pi interaction is shown as solid line.

residue Phe538. This interaction kept 2-*O*-caffeoyl tartaric acid close to key binding site residue Gln535, thereby hindering binding of natural substrate to PDE4D (Fig. 3(a)). Moreover, control RS-25344 had a hydrogen bond with Gln535, which further supported the importance of blocking substrate from interacting with this residue (Fig. 3(d)). The second ranking TCM compound, mumefural, did not share pi-pi interaction as observed in other compounds. However, mumefural had a hydrogen bonding with catalytic residue His 326 (Fig. 3(b)). This hydrogen bonding interaction, which completely blocks catalytic process, suggests an alternate strategy of inhibiting PDE4D, in addition to mechanism observed with 2-*O*-caffeoyl tartaric acid.

### Molecular dynamics simulation

After docking simulation, we performed molecular dynamics simulation to further analyze the interactions between protein and ligand. Root mean squared deviation (RMSD) was calculated to study atomic fluctuations during MD simulation. Fig. 4 shows the whole molecule RMSD values for the control-PDE4D complexes and the two TCM ligand-PDE4D complexes. The control-PDE4D complexes stabilized after 10 ns at 1.4 Å while the 2-*O*-caffeoyl tartaric acid complex stabilized after 14 ns. The mumefural-protein complex, in contrast, appears to remain fluctuating even after 20 ns of simulation run. However, comparison of binding site snapshots taken at different time period showed that the position of the binding site residues were

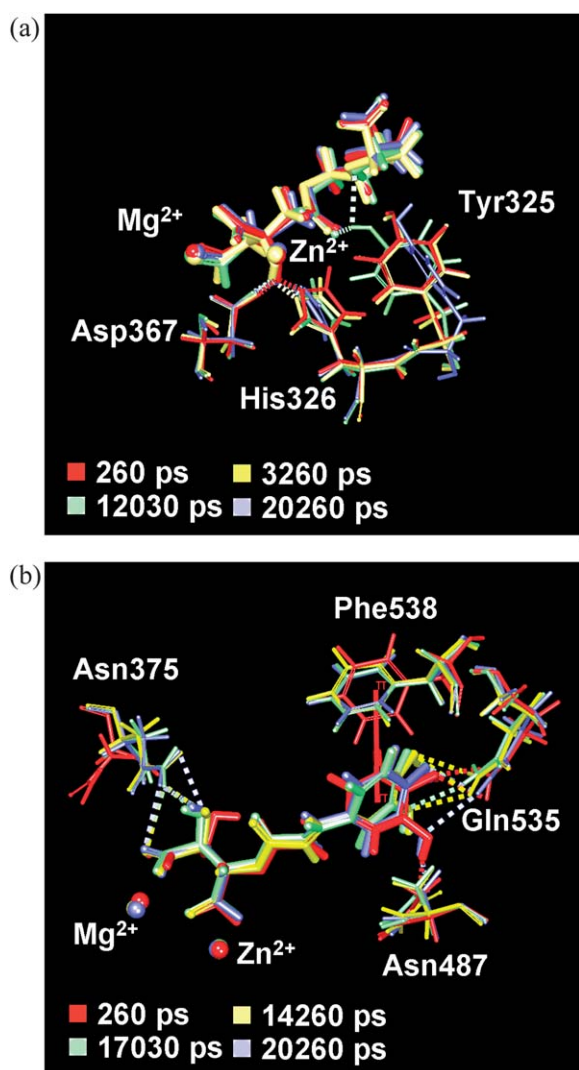


**Fig. 4** RMSD of PDE4D in complex with 2-*O*-caffeoyl tartaric acid, mumefural, cAMP, and RS-25344 respectively (top), and ligand RMSD of 2-*O*-caffeoyl tartaric acid, mumefural, cAMP, and RS-25344 respectively (bottom).

relatively unchanged (Fig. 5(a)) and that all receptor-ligand interaction were conserved during simulation run. For these reasons, we did not continue production run for mumefural-PDE4D complex after 20 ns.

As for ligand RMSD, all ligands, except 2-*O*-caffeoyl tartaric acid, stabilized after 3 ns (Fig. 4). Ligand 2-*O*-caffeoyl tartaric acid had dramatic atom fluctuations from 15 ns to 17 ns.





However, as shown in Fig. 5(b), this large atomic fluctuation was due to intrinsic rotation of benzene substructure and did not influence the existed electrostatic interactions and hydrogen bonds.

Shown in Fig. 6 are snapshots of different receptor-ligand complexes taken at the end of molecular dynamics simulation. After 20 ns of simulation, only RS-25344 maintained the pi-pi interaction to Phe538 (Fig. 6(a)); all other receptor-ligand pi-pi interactions diminished (Video S1, Video S2). RS-25344 was co-crystallized with PDE4D and therefore, could induce protein conformation change prior to crystallography process to accommodate for binding.

In contrast to pi-pi interaction, electrostatic interactions of ligands to binding site residues, Asp367 and Asp484, through divalent metal cations, were still retained after MD simulation run (Fig. 6). Interactions of Zn<sup>2+</sup> to Asp379 and Asp484 as well as interaction of Mg<sup>2+</sup> to Asp484 were conserved throughout the MD simulation run (Video S1, Video S2). Furthermore, as shown in Fig. 7, with the exception of RS-25344 to Mg<sup>2+</sup>,

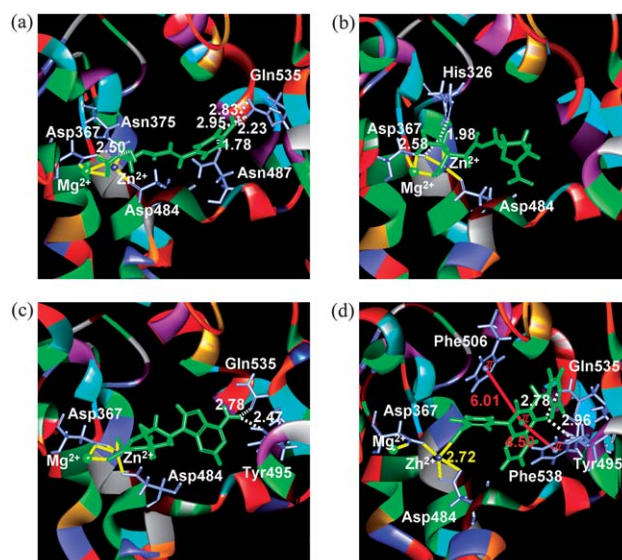
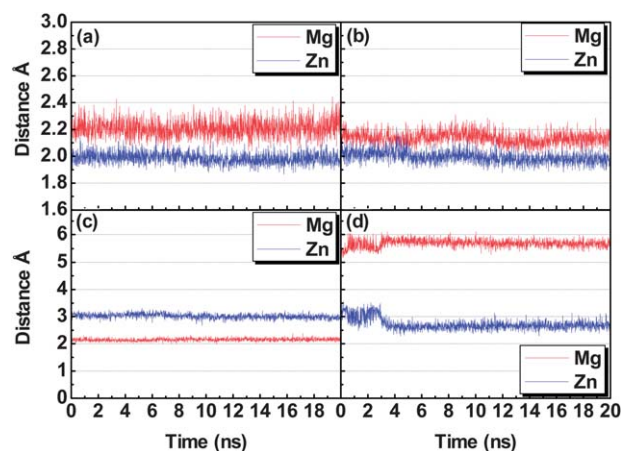


Fig. 7 Distances of electrostatic bonds between Mg<sup>2+</sup>, Zn<sup>2+</sup> and (a) 2-*O*-caffeoyl tartaric acid, (b) mumeufural, (c) cAMP and (d) RS-25344.



electrostatic interactions of ligands to both cations were stabilized and have only minor fluctuations.

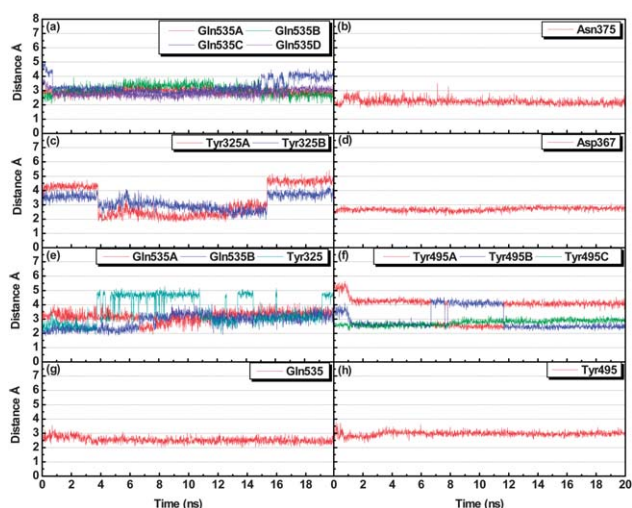
For hydrogen bond interactions, the average length and occupancy during the simulation run are summarized in Table 2.

The lengths of different hydrogen bonds with respect to time are shown in Fig. 8. Although from analyzing ligand docking poses, there was no interaction between cAMP or 2-*O*-caffeoyl tartaric acid to Gln535, this interaction appears during the MD simulation for these two compounds. With the exception of mumeufural, all compounds had hydrogen bond interaction to Gln535 (Fig. 8) with an average distance of 3.0 Å. Although mumeufural did not have hydrogen bonding with Gln535, it established stable, continuous hydrogen bonding with the key active site residue His326 (Table 2). In addition to ligand hydrogen bonding to

**Table 2** H-bond interactions of PDE4D with 2-*O*-caffeoyl tartaric acid, mumeifural, cAMP, and RS-25344<sup>a</sup>

Ligand	H-bond	Ligand Atom	Amino acid	Max. distance	Min. distance	Average distance	H-bond occupancy
2- <i>O</i> -Caffeoyl tartaric acid	1	O6	Asn375:HD22	5.94	1.70	2.24	91.36%
	2	O28	Gln535:HE21	3.64	2.29	2.92	1.05%
	3	O28	Gln535:HE22	3.99	2.08	3.03	6.24%
	4	O31	Gln535:HE21	5.05	2.49	3.31	0.15%
	5	O31	Gln535:HE22	4.10	2.14	2.83	8.84%
Mumeifural	1	O13	TYR325:HH	5.42	1.75	3.30	37.06%
	2	O17	TYR325:HH	4.34	1.92	3.20	8.39%
	3	H10	HIS326:NE2	2.63	1.73	1.99	99.85%
	4	H10	ASP367:OD2	3.11	2.20	2.68	11.24%
cAMP	1	O15	TYR325:HH	5.28	1.94	3.63	11.74%
	2	N8	Tyr495:HH	3.29	2.16	2.75	10.84%
	3	H29	Tyr495:OH	5.67	2.20	3.81	11.79%
	4	H30	Tyr495:OH	4.54	2.15	2.96	34.17%
	5	H29	Gln535:OE1	4.20	1.90	3.20	5.99%
	6	H30	Gln535:OE1	4.04	1.79	2.86	29.72%
RS-25344	1	O21	Tyr495:HH	3.80	2.31	2.97	0.55%
	2	O21	Gln535:HE22	3.56	2.01	2.53	46.15%

<sup>a</sup> H-bond occupancy cutoff: 2.5 Å.



**Fig. 8** Distances of hydrogen bonds between 2-*O*-caffeoyl tartaric acid and (a) Gln535 or (b) Asn375. Distances of hydrogen bonds between mumeifural and (c) Tyr325 or (d) Asp367, Distances of hydrogen bond between cAMP and (e) Gln535 or Tyr325 or (f) Tyr495. Distances of hydrogen bond between RS-25344 and (g) Gln535, (h) Tyr495.

Gln535 and His326, several hydrogen bonding were observed between ligands and Tyr325, Asp367 and Asn375 (Table 2). Fig. 8 shows that both mumeifural and cAMP had fluctuated hydrogen bond interactions with Tyr325 (Video S1) which could hold ligand close to active site His326. Both controls formed stable hydrogen bond interactions with Tyr495. In addition, 2-*O*-caffeoyl tartaric acid formed strong hydrogen bond interaction with Asn375 after a short period of molecular dynamics simulation. These interactions were proposed to further stabilize the two TCM ligands in PDE4D binding pocket.

### Genetic function approximation, multiple linear regression and support vector machine analysis

To select the most representative descriptors for building QSAR, genetic function approximation was used. The selected

descriptors were Dipole mag, Dipole Y, Jurs FNSA, Jurs PPSA3, Jurs RASA, Jurs RNCG, Jurs TASA, Jurs WPSA3, shadow X length and shadow Y length. The values of descriptors for the training set and the test set were shown in Supplementary Table S1.

These selected descriptors can be broadly divided into two groups: the electronic descriptors and the spatial descriptors. The spatial descriptors include Dipole mag (Dipole magnitude) and Dipole Y (Dipole moment in Y dimension), which are both dipole moment descriptors computed from partial atomic charges and atomic coordinates.<sup>45</sup> In our case, the numerical values of the training set and the test were in the range of 0.55285 to 7.37706 for Dipole mag and in the range of -4.25548 to 6.024515 for Dipole Y. Most of the spatial descriptors are Jurs descriptors, which is dependant on atomic partial charges and solvent-accessible surface areas of individual atoms under investigation.<sup>46</sup> The Jurs RASA and Jurs TASA are related to hydrophobic surface area whereas Jurs RNCG is related to negative charge. The other three Jurs descriptors, Jurs FNSA3, Jurs PPSA3 and Jurs WPSA3, are significantly dependent on partial charge states. Both Shadow Xlength and Shadow Ylength are shadow indices. These two descriptors stand for length of molecule projected on X and Y dimension respectively.<sup>47</sup>

With these ten descriptors selected by GFA, an MLR model was built using a training set of 51 compounds. A test set of 11 compounds was used to evaluate the built model. The following linear model was obtained:

$$\begin{aligned} \text{predicted\_IC}_{50} = & -23.43 + 0.1383 \times \text{Dipole\_mag} + 0.0921 \\ & \times \text{Dipole\_Y} - 31.708 \times \text{Jurs\_FNSA3} - 0.2866 \times \text{Jurs\_PPSA3} \\ & + 28.543 \times \text{Jurs\_RASA} + 20.607 \times \text{Jurs\_RNCG} - 0.0330 \\ & \times \text{Jurs\_TASA} + 0.691 \times \text{Jurs\_WPSA3} + 0.428 \\ & \times \text{Shadow\_Xlength} + 0.528 \times \text{Shadow\_Ylength} \end{aligned}$$

The same training set and test set were used to build and to test the SVM model. The statistical significances of the MLR and the SVM model are shown in Fig. 9. The training set square

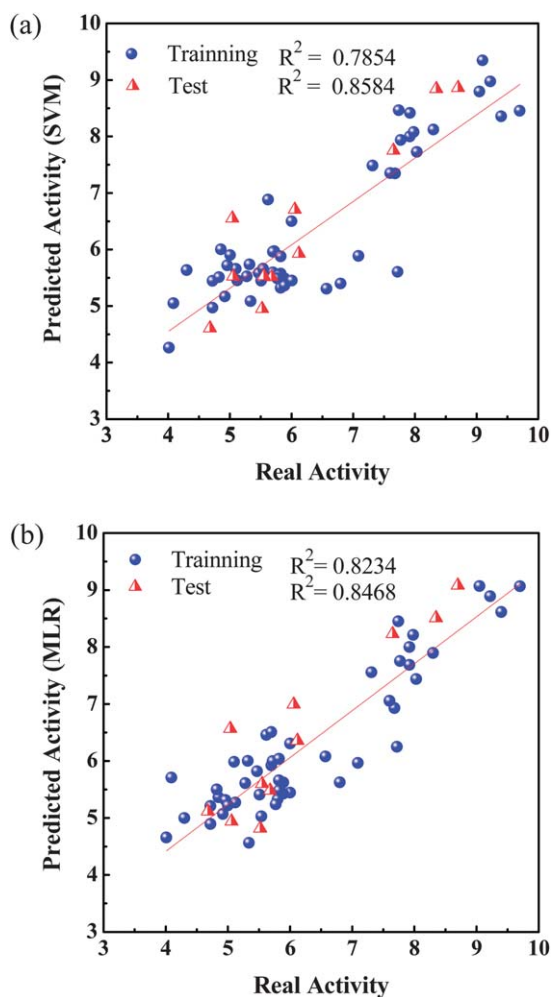


Fig. 9 The predicted models: (a) MLR and (b) SVM.

correlation coefficients ( $R^2$ ) were 0.8234 for MLR and 0.7854 for SVM. The test set  $R^2$  were 0.8468 for MLR and 0.8584 for SVM.

Both models built by MLR and SVM predicted reasonable bioactivities for the TCM ligands. The MLR model gives predicted  $pIC_{50}$  values of 10.239 and 9.0159 for 2-*O*-caffeoyl tartaric acid and mumefural, respectively (Table 3). The SVM model, on the other hand, gives predicted  $pIC_{50}$  values of 6.236 and 5.758 for top 1 and top2, respectively (Table 3). These predicted activities are all better than the control cAMP and RS-25344. However, since the predicted  $pIC_{50}$  values obtained from MLR were based on linear extrapolation, the SVM model, which was based on non-linear function, should yield a more accurate prediction. Hence, the  $pIC_{50}$  values from SVM models should be

**Table 3** Activity prediction of 2-*O*-caffeoyl tartaric acid, mumefural, cAMP, and RS-25344 by SVM and MLR models

Predicted model	2- <i>O</i> -Caffeoyl_			
	Cyclic_AMP	RS-25344	tartaric_acid	Mumefural
SVM	5.571	5.433	6.236	5.758
MLR	8.305	6.454	10.239	9.016

a better representation of the two TCM compound bioactivities. Nevertheless, these results further support 2-*O*-caffeoyl tartaric acid and mumefural as potent PDE4D inhibitors.

## Conclusion

PDE4D has been found as a stroke-related gene previously in genome-wide association study. In this study, we report several binding patterns observed from docking of TCM and natural compounds and molecular dynamics simulation. We observed that the divalent metal cations act as a linkage between ligand and two of key residues, Asp367 and Asp484. This metal ion dependency called for negative charged substructure for stable interaction. In our case, all top TCM ligands have carboxylate groups which formed stable electrostatic interactions. The ligand hydrogen bonding interactions to key residue Gln535 or His326 could sufficiently hinder binding of natural substrate, thereby inhibiting PDE4D function. Furthermore, hydrogen bonding interaction between ligands to Tyr325, Asp367 and Asn375 were suggested to strengthen and stabilize ligand inside PDE4D binding site. To predict the bioactivity of the TCM ligands, we also built MLR and SVM models. The top two TCM compounds had reasonable  $pIC_{50}$  values. Overall, we suggest the above mentioned binding characteristics could be important for ligand inhibitory activity. Hence, we propose the top two TCM candidates, 2-*O*-caffeoyl tartaric acid and mumefural, as potential PDE4D inhibitors.

## Acknowledgements

The research was supported by grants from the National Science Council of Taiwan (NSC 99-2221-E-039-013-), China Medical University (CMU98-TCM, CMU99-TCM, CMU99-S-02) and Asia University (CMU98-ASIA-09). This study is also supported in part by Taiwan Department of Health Clinical Trial and Research Center of Excellence (DOH100-TD-B-111-004) and Taiwan Department of Health Cancer Research Center of Excellence (DOH100-TD-C-111-005). We are grateful to the National Center of High-performance Computing for computer time and facilities.

## References

- 1 C. Y. C. Chen, *PLoS One*, 2011, **6**, e15939.
- 2 D. Lloyd-Jones, R. J. Adams, T. M. Brown, M. Carnethon, S. Dai, G. De Simone, T. B. Ferguson, E. Ford, K. Furie, C. Gillespie, A. Go, K. Greenlund, N. Haase, S. Hailpern, P. M. Ho, V. Howard, B. Kissela, S. Kittner, D. Lackland, L. Lisabeth, A. Marelli, M. M. McDermott, J. Meigs, D. Mozaffarian, M. Mussolino, G. Nichol, V. L. Roger, W. Rosamond, R. Sacco, P. Sorlie, T. Thom, S. Wasserthiel-Smoller, N. D. Wong and J. Wylie-Rosett, *Circulation*, 2010, **121**, e46–e215.
- 3 M. N. Beckley, *Am. J. Occup. Ther.*, 2006, **60**, 129–135.
- 4 S. Gretarsdottir, G. Thorleifsson, S. T. Reynisdottir, A. Manolescu, S. Jonsdottir, T. Jonsdottir, T. Gudmundsdottir, S. M. Bjarnadottir, O. B. Einarsson, H. M. Gudjonsdottir, M. Hawkins, G. Gudmundsson, H. Gudmundsdottir, H. Andrason, A. S. Gudmundsdottir, M. Sigurdardottir, T. T. Chou, J. Nahmias, S. Goss, S. Sveinbjornsdottir, E. M. Valdimarsson, F. Jakobsson, U. Agnarsson, V. Gudnason, G. Thorgeirsson, J. Fingerle, M. Gurney, D. Gudbjartsson, M. L. Frigge, A. Kong, K. Stefansson and J. R. Gulcher, *Nat. Genet.*, 2003, **35**, 131–138.

- 5 J. M. Staton, M. S. Sayer, G. J. Hankey, J. Attia, A. Thakkinstian, Q. Yi, V. J. Cole, R. Baker and J. W. Eikelboom, *J. Neurol., Neurosurg. Psychiatry*, 2006, **77**, 1067–1069.
- 6 H. Xue, H. Wang, X. D. Song, W. J. Li, K. Sun, W. L. Zhang, X. J. Wang, Y. B. Wang and R. T. Hui, *Clin. Sci.*, 2009, **116**, 335–340.
- 7 Y. Sun, Y. Y. Huang, X. Chen, Y. Liu, X. Z. Lu, Y. Y. Shi, W. Tang, J. D. Yang, W. Y. Chen, X. Z. Zhao, L. H. Gao, S. Li, G. Y. Feng and L. He, *Clin. Sci.*, 2009, **117**, 265–272.
- 8 T. Nakayama, S. Asai, N. Sato and M. Soma, *Stroke*, 2006, **37**, 69–76.
- 9 M. J. van Rijn, A. J. Slooter, A. F. Schut, A. Isaacs, Y. S. Aulchenko, P. J. Snijders, L. J. Kappelle, J. C. van Swieten, B. A. Oostra and C. M. van Duijn, *Neurology*, 2005, **65**, 1203–1209.
- 10 D. Saleheen, S. Bukhari, S. R. Haider, A. Nazir, S. Khanum, S. Shafqat, M. K. Anis and P. Frossard, *Stroke*, 2005, **36**, 2275–2277.
- 11 S. Nilsson-Ardnor, P. G. Wiklund, P. Lindgren, A. K. Nilsson, T. Janunger, S. A. Escher, B. Hallbeck, B. Stegmayr, K. Asplund and D. Holmberg, *Stroke*, 2005, **36**, 1666–1671.
- 12 S. Bevan, L. Porteous, M. Sitzer and H. S. Markus, *Stroke*, 2005, **36**, 949–953.
- 13 V. H. Brophy, S. K. Ro, B. K. Rhees, L. Y. Lui, J. M. Lee, N. Umblas, L. G. Bentley, J. Li, S. Cheng, W. S. Browner and H. A. Erlich, *Stroke*, 2006, **37**, 1385–1390.
- 14 H. P. Lu, Y. J. Chie, M. S. Yang, C. S. Lee, J. J. Fu, J. S. Yang, T. W. Tan, S. H. Wu, Y. S. Ma, S. W. Ip and J. G. Chung, *Int. J. Oncol.*, 2010, **36**, 1477–1484.
- 15 M. L. Lin, Y. C. Lu, J. G. Chung, Y. C. Li, S. G. Wang, S. H. Ng, C. Y. Wu, H. L. Su and S. S. Chen, *Cancer Lett.*, 2010, **291**, 46–58.
- 16 C. L. Chang, L. J. Zhang, R. Y. Chen, C. C. Wu, H. C. Huang, M. C. Roy, J. P. Huang, Y. C. Wu and Y. H. Kuo, *Bioorg. Med. Chem.*, 2010, **18**, 518–525.
- 17 C. Y. Liew, X. H. Ma, X. H. Liu and C. W. Yap, *J. Chem. Inf. Model.*, 2009, **49**, 877–885.
- 18 C. Y. Chen and C. Y. C. Chen, *J. Mol. Graphics Modell.*, 2010, **29**, 21–31.
- 19 H. J. Huang, C. Y. Chen, H. Y. Chen, F. J. Tsai and C. Y. C. Chen, *J. Taiwan Inst. Chem. Eng.*, 2010, **41**, 352–359.
- 20 C. Y. C. Chen, *J. Mol. Graphics Modell.*, 2009, **28**, 261–269.
- 21 T. T. Chang, H. J. Huang, K. J. Lee, H. W. Yu, H. Y. Chen, F. J. Tsai, M. F. Sun and C. Y. C. Chen, *J. Biomol. Struct. Dyn.*, 2010, **28**, 309–321.
- 22 H. J. Huang, H. W. Yu, C. Y. Chen, C. H. Hsu, H. Y. Chen, K. J. Lee, F. J. Tsai and C. Y. C. Chen, *J. Taiwan Inst. Chem. Eng.*, 2010, **41**, 623–635.
- 23 H. J. Huang, K. J. Lee, H. W. Yu, H. Y. Chen, F. J. Tsai and C. Y. C. Chen, *J. Biomol. Struct. Dyn.*, 2010, **28**, 187–200.
- 24 H. J. Huang, K. J. Lee, H. W. Yu, C. Y. Chen, C. H. Hsu, H. Y. Chen, F. J. Tsai and C. Y. C. Chen, *J. Biomol. Struct. Dyn.*, 2010, **28**, 23–37.
- 25 C. Y. C. Chen, *J. Biomol. Struct. Dyn.*, 2010, **27**, 627–640.
- 26 C. Y. C. Chen, *J. Taiwan Inst. Chem. Eng.*, 2010, **41**, 143–149.
- 27 C. Y. Chen, H. J. Huang, F. J. Tsai and C. Y. C. Chen, *J. Taiwan Inst. Chem. Eng.*, 2010, **41**, 8–15.
- 28 A. B. Burgin, O. T. Magnusson, J. Singh, P. Witte, B. L. Staker, J. M. Bjornsson, M. Thorsteinsdottir, S. Hrafnisdottir, T. Hagen, A. S. Kiselyov, L. J. Stewart and M. E. Gurney, *Nat. Biotechnol.*, 2010, **28**, 63–70.
- 29 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
- 30 C. M. Venkatachalam, X. Jiang, T. Oldfield and M. Waldman, *J. Mol. Graphics Modell.*, 2003, **21**, 289–307.
- 31 D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel and S. T. Freer, *Chem. Biol.*, 1995, **2**, 317–324.
- 32 I. Muegge and Y. C. Martin, *J. Med. Chem.*, 1999, **42**, 791–804.
- 33 M. Avriel, *Nonlinear Programming: Analysis and Methods*, Press: Dover Publishing, 2003.
- 34 R. Fletcher and C. M. Reeves, *Comput. J.*, 1964, **7**, 149–154.
- 35 <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- 36 G. L. Card, L. Blasdel, B. P. England, C. Zhang, Y. Suzuki, S. Gillette, D. Fong, P. N. Ibrahim, D. R. Artis, G. Bollag, M. V. Milburn, S. H. Kim, J. Schlessinger and K. Y. Zhang, *Nat. Biotechnol.*, 2005, **23**, 201–207.
- 37 R. Aspiotis, D. Deschenes, D. Dube, Y. Girard, Z. Huang, F. Laliberte, S. Liu, R. Papp, D. W. Nicholson and R. N. Young, *Bioorg. Med. Chem. Lett.*, 2010, **20**, 5502–5505.
- 38 K. Naganuma, A. Omura, N. Maekawara, M. Saitoh, N. Ohkawa, T. Kubota, H. Nagumo, T. Kodama, M. Takemura, Y. Ohtsuka, J. Nakamura, R. Tsujita, K. Kawasaki, H. Yokoi and M. Kawanishi, *Bioorg. Med. Chem. Lett.*, 2009, **19**, 3174–3176.
- 39 D. Rogers and A. J. Hopfinger, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 854.
- 40 V. N. Vapnik, *IEEE Trans. Neural Networks*, 1999, **10**, 988–999.
- 41 C. Cortes and V. Vapnik, *Machine Learning*, 1995, **20**, 273–297.
- 42 M. Goodarzi, M. P. Freitas and R. Jensen, *Chemom. Intell. Lab. Syst.*, 2009, **98**, 123–129.
- 43 S. Riahi, E. Pourbasheer, M. R. Ganjali and P. Norouzi, *Chem. Biol. Drug Des.*, 2009, **73**, 558–571.
- 44 A. J. Smola and B. Scholkopf, *Stat. Comput.*, 2004, **14**, 199–222.
- 45 A. J. Hopfinger, *Conformational properties of macromolecules*, Academic Press, New York, 1973.
- 46 D. T. Stanton and P. C. Jurs, *Anal. Chem.*, 1990, **62**, 2323–2329.
- 47 R. H. Rohrbaugh and P. C. Jurs, *Anal. Chem.*, 1987, **59**, 1048–1054.