

2D Autocorrelation, CoMFA, and CoMSIA modeling of protein tyrosine kinases' inhibition by substituted pyrido[2,3-*d*]pyrimidine derivatives

Julio Caballero,^{a,*} Michael Fernández,^b Mario Saavedra^a
and Fernando D. González-Nilo^a

^a*Centro de Bioinformática y Simulación Molecular, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile*

^b*Molecular Modeling Group, Center for Biotechnological Studies, University of Matanzas, Matanzas, Cuba*

Received 19 July 2007; revised 21 September 2007; accepted 10 October 2007

Available online 13 October 2007

Abstract—2D Autocorrelation, comparative molecular field analysis (CoMFA), and comparative molecular similarity indices analysis (CoMSIA) were undertaken for a series of substituted pyrido[2,3-*d*]pyrimidine derivatives to correlate platelet-derived growth factor receptor (PDGFR), fibroblast growth factor receptor (FGFR), and c-Src tyrosine kinases' inhibition with 2D and 3D structural properties of 22 known compounds. QSAR models with considerable internal as well as external predictive ability were obtained. The relevant 2D autocorrelation descriptors for modeling each protein tyrosine kinase (PTK) inhibitory activity were selected by genetic algorithm (GA) and multiple linear regression (MLR) approach. The 2D autocorrelation space brings different descriptors for each PTK inhibition and suggests the atomic properties relevant for the inhibitors to interact with each PTK active site. CoMFA and CoMSIA were developed with a focus on interpretative ability using coefficient contour maps. CoMSIA produced significantly better results for all correlations. The results indicate a strong correlation between the inhibitory activity of the modeled compounds and the hydrophobic and H-bond donor fields around them.

© 2007 Elsevier Ltd. All rights reserved.

1. Introduction

Protein tyrosine kinases (PTKs) catalyze transfer of the γ -phosphate group of ATP to the hydroxyl group of tyrosine residues of proteins. PTKs regulate signaling pathways for a broad spectrum of cellular processes including responses to growth factors, neurotransmitters, and hormones, activation of the immune response, regulation of cell–cell and cell–extracellular matrix interactions, as well as development, oncogenesis, and angiogenesis.^{1,2} The aberrant over- or underexpression of PTKs or mutations that induce alterations in the regulation of PTK activity can lead directly to the perturbation of any of these processes. In this sense, the development of specific PTK inhibitors (PTKIs) as pharmacological tools and potential antiproliferative

agents has therefore become an active area of research.^{3,4}

Actually, several PTKIs have been approved by the United States Food and Drug Administration (FDA) including small molecules such as ZD-1839 (Iressa) against intracellular domain of endothelial growth factor receptor (EGFR),⁵ STI-571 (Gleevec) against Abl,⁶ and monoclonal antibodies such as Herceptin (Trastuzumab) against HER2⁷ and Bevacizumab (Avastin) against vascular endothelial growth factor receptor (VEGFR).⁸ For clinical applications, PTKIs must show high cellular and in vivo potency and must be selective. Since most PTKIs that have been tested in clinical trials act by directly competing with ATP at the ATP-binding site of the PTK, the selectivity is an essential requirement for clinically effective PTKIs. For this reason, an understanding of the structural differences among the ATP-binding sites of PTKs is of crucial importance.

Computational models that are able to predict PTK structures^{9,10} and the biological activities of PTK-binding drugs by their structural properties are powerful

Keywords: Protein tyrosine kinase inhibitors; QSAR analysis; 2D Autocorrelation space; CoMFA; CoMSIA.

* Corresponding author. Tel.: +56 71 201 662; fax: +56 71 201 561; e-mail addresses: jcaballero@utalca.cl; jmc77@yahoo.com

tools to design highly active molecules. In this sense, quantitative structure–activity relationship (QSAR) studies have been successfully applied for modeling biological activities of natural and synthetic chemicals.¹¹ Much experimental work has been carried out on this family of proteins, and a number of QSAR studies have been carried out as well. Novič et al.¹² applied an artificial neural network (ANN) based on a counterpropagation algorithm to develop QSARs in a large dataset of 105 flavonoid derivatives that inhibit the enzyme p56^{lck} PTK using both classical and quantum-chemical descriptors. Kurup et al.¹³ conducted an extensive review of comparative QSAR models for tyrosine kinase inhibitors. All the models reported were linear in nature and were developed using a limited number of descriptors. Shen et al.¹⁴ developed a series of linear regression models for the set of 1-phenylbenzimidazoles using electronic descriptors and a partial least square (PLS) routine to build the final models. Guha and Jurs¹⁵ modeled the platelet-derived growth factor receptor tyrosine kinase (PDGFR-TK) inhibition of 79 piperazinylquinazoline analogues using PLS and ANNs. Authors found that ANN model was more predictive, but non-interpretable. The most important feature for the relationship was the hydrophobic surface area. Chen et al.¹⁶ carried out a QSAR study that covers different kinds of inhibitors of EGFR tyrosine kinase. Authors combined molecular docking and 3D-QSAR to study 124 reported inhibitors with different scaffolds.

In our current work, we carried out 2D- and 3D-QSAR studies for establishing correlations between the structural properties of 22 substituted pyrido[2,3-*d*]pyrimidine derivatives (SPPs)¹⁷ (the chemical structures are shown in Table 1) and their inhibitory activities against PDGFR, fibroblast growth factor receptor (FGFR), and c-Src tyrosine kinases (PDGFR-TK, FGFR-TK, and c-Src-TK). 2D-QSAR analysis was performed using multiple linear regression (MLR), genetic algorithm (GA) and 2D autocorrelation vectors. 3D-QSAR analysis was performed using comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA).

2. Results and discussion

The studied dataset includes compounds which are potent inhibitors of PDGFR-TK, FGFR-TK, and c-Src-TK. Correlation matrix (Table 2) shows that inhibitory activities of SPPs employed in this study against PTKs are related to each other. In fact, the activities against PDGFR-TK and c-Src-TK are highly related.

2.1. 2D Autocorrelation approach

Three 2D autocorrelation MLR models are reported in this work. In total, nine descriptors from the whole 2D autocorrelation pool were employed. The colinearity of the variables should be as low as possible for guarantying the absence of redundant information.¹⁸ The correlation of each one of these descriptors in these

equations with each other was calculated (Table 3). There are only two correlated pairs ($R^2 > 0.7$) from 36 pairs. Furthermore, there are two pairs with correlations between 0.5 and 0.7, 14 pairs with correlations between 0.1 and 0.5, and 18 pairs with correlations between 0 and 0.1.

MLR models for inhibitory activities of SPPs against three PTKs were obtained with acceptable statistical significances and predictive power (Eqs. 1–3).

2DA-PDGFR-TK:

$$\begin{aligned} \log(10^6/\text{IC}_{50}) = & -24.042 \times \text{MATS1m} - 3.213 \\ & \times \text{MATS8v} - 9.196 \times \text{MATS6e} \\ & + 25.481 \end{aligned} \quad (1)$$

$$\begin{aligned} N = 18, \quad R^2 = 0.716, \quad S = 0.634, \quad p < 10^{-5} \\ Q^2 = 0.598, \quad S_{\text{CV}} = 0.672 \end{aligned}$$

2DA-FGFR-TK:

$$\begin{aligned} \log(10^6/\text{IC}_{50}) = & 36.021 \times \text{MATS6m} - 20.634 \\ & \times \text{MATS1e} - 6.513 \times \text{GATS5v} \\ & - 26.908 \end{aligned} \quad (2)$$

$$\begin{aligned} N = 18, \quad R^2 = 0.808, \quad S = 0.561, \quad p < 10^{-5} \\ Q^2 = 0.676, \quad S_{\text{CV}} = 0.653 \end{aligned}$$

2DA-c-Src-TK:

$$\begin{aligned} \log(10^6/\text{IC}_{50}) = & 0.160 \times \text{ATS1v} + 2.262 \\ & \times \text{MATS3v} - 7.634 \times \text{GATS1v} \\ & + 5.328 \end{aligned} \quad (3)$$

$$\begin{aligned} N = 18, \quad R^2 = 0.860, \quad S = 0.654, \quad p < 10^{-5} \\ Q^2 = 0.772, \quad S_{\text{CV}} = 0.739 \end{aligned}$$

In Eqs. 1–3, N is the number of compounds included in the training set, R^2 is the square of correlation coefficients, S is the standard deviation of the regression, p is the significance of the variables in the model, and Q^2 and S_{CV} are the correlation coefficients and standard deviations of the leave-one-out (LOO) cross-validation, respectively.

The MLR training set predictions ($\log(10^6/\text{IC}_{50})$) for the SPPs against all three PTKs appear in Table 1. In turn, plots of training set and LOO cross-validation predictions versus experimental $\log(10^6/\text{IC}_{50})$ values for the MLR models are shown in Figure 1. In general, MLR models were able to explain data variance and were quite stable to the inclusion–exclusion of compounds as measured by LOO correlation coefficients ($Q^2 > 0.5$).

Table 1. Experimental and predicted PTK inhibitory activities of substituted pyrido[2,3-*d*]pyrimidine derivatives

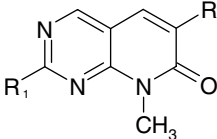
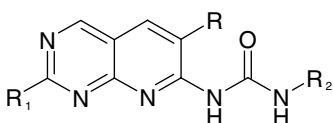
<div><div></div><div>1-9, 18-21</div><div></div><div>10-17, 22</div></div>												
Compound	R	R ₁	R ₂	PDGFR-TK			FGFR-TK			c-Src-TK		
				Exp.	Predicted		Exp.	Predicted		Exp.	Predicted	
					2DA-PDGFR-TK	CoMSIA-HD-1		2DA-FGFR-TK	CoMSIA-HD-2		2DA-c-Src-TK	CoMSIA-HD-3
<i>Training set</i>												
1	Ph(2,6-diCl)	NH ₂	—	2.31	2.00	1.94	2.88	3.29	2.65	3.59	2.95	3.08
2	Ph(2,6-diCl)	H	—	1.30	1.82	1.13	1.30	1.36	1.23	2.00	2.22	1.60
3	Ph(2,6-diCl)	Me	—	1.30	1.94	1.21	1.30	2.09	1.36	1.30	2.48	1.37
4	Ph(2,6-diCl)	NMe ₂	—	1.30	1.18	1.24	1.30	1.00	1.38	1.30	1.78	1.62
5	Ph(2,6-diCl)	NH(CH ₂) ₃ -N-Me-piperazine	—	2.57	2.66	2.78	3.46	3.57	3.30	3.63	3.08	3.61
6	Ph(2,6-diCl)	NH(CH ₂) ₃ NEt ₂	—	2.05	2.09	2.71	2.71	3.32	2.79	3.02	3.02	3.29
7	Ph(2,6-diCl)	NHPh(3-O(CH ₂) ₂ NEt ₂)	—	3.99	3.05	3.81	4.32	3.79	4.41	5.05	4.34	5.01
8	Ph(2,6-diCl)	NHPh(4-CH ₂ CO ₂ H)	—	4.15	3.90	4.08	4.21	3.87	4.33	5.00	4.38	5.14
9	Ph(2,6-diCl)	NHPh(4-piperazine-NMe)	—	3.94	2.98	3.82	4.29	3.54	4.31	5.00	4.33	5.05
10	Ph(2,6-diCl)	NH(CH ₂) ₄ NEt ₂	<i>t</i> -Bu	3.44	2.60	3.30	4.32	4.13	4.08	4.64	3.86	4.31
11	Ph(2,6-diCl)	NH(CH ₂) ₃ -N-Me-piperazine	Cyclohexyl	3.58	3.65	3.32	4.40	3.94	4.43	4.51	4.82	4.68
12	Ph(2,6-diCl)	NH(CH ₂) ₃ -N-Me-piperazine	Ph-4-OCH ₃	3.10	2.99	2.98	3.86	4.40	3.82	4.55	5.10	4.61
13	Ph(2,6-diCl)	NH(CH ₂) ₃ -N-Me-piperazine	Benzyl	2.70	3.11	2.92	3.90	4.17	4.05	4.52	5.26	4.66
14	Ph(2,6-diCl)	NHPh(4-O(CH ₂) ₂ NEt ₂)	<i>t</i> -Bu	2.96	3.87	3.25	4.29	4.39	4.25	4.62	5.06	4.46
15	Ph(3,5-diNMe ₂)	NH ₂	<i>t</i> -Bu	1.30	1.33	1.44	1.80	2.35	2.63	1.30	0.94	1.48
16	Ph(3,5-diOEt)	NH ₂	<i>t</i> -Bu	1.30	1.43	1.36	2.78	2.28	2.97	1.30	1.73	1.19
17	Ph(3,5-diOMe)	NH ₂	<i>t</i> -Bu	1.30	0.95	1.38	4.22	3.35	3.08	1.30	0.96	1.24
18	Ph(3,5-diOMe)	NH ₂	—	1.37	2.41	1.33	3.90	4.39	4.17	1.30	1.62	1.54
<i>Test set</i>												
19	Ph(2,6-diCl)	NHPh	—	3.48	3.42	3.81	3.29	3.00	4.28	4.62	3.88	4.85
20	Ph(2,6-diCl)	NH(CH ₂) ₃ -morpholine	—	2.20	2.72	2.77	2.90	3.41	3.29	3.28	3.11	3.61
21	Ph(2,6-diCl)	NHPh(4-O(CH ₂) ₂ NEt ₂)	—	4.02	4.05	3.74	4.36	3.82	4.22	5.05	4.34	4.91
22	Ph(3,5-diCF ₃)	NH ₂	<i>t</i> -Bu	1.30	1.80	1.66	1.30	1.68	2.12	1.30	1.50	2.14

Table 2. Correlation matrix for the substituted pyrido[2,3-*d*]pyrimidine activities against PTKs

	PDGFR-TK	FGFR-TK	c-Src-TK
PDGFR-TK	1		
FGFR-TK	0.565	1	
c-Src-TK	0.925	0.537	1

2DA-PDGFR-TK and 2DA-FGFR-TK models included atomic mass, van der Waals volume, and Sanderson electronegativity weighted terms, with a negative influence of van der Waals volume in both models. By contrast, 2DA-c-Src-TK model only included van der Waals volume weighted terms. The computation of 2D autocorrelation descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and leads to different autocorrelation vectors corresponding to the lengths of the structural fragments. Bearing in mind this aspect, the interpretation of 2D autocorrelation descriptors is uneasy. Nevertheless, these models have the potential of predicting new compounds and discriminating their effect against the modeled PTKs.

2.2. CoMFA and CoMSIA results

Figure 2 shows the aligned molecules within the grid box (grid spacing 2.0 Å) used to generate the CoMFA and CoMSIA columns. The stepwise development of CoMFA and CoMSIA models using different fields is presented in Table 4. The predictability of the models is the most important criterion for assessment of both methods. The best CoMFA models describing PDGFR-TK and c-Src-TK inhibition used steric field ($Q^2 = 0.614$ and 0.724 , respectively). Instead, all CoMFA models describing FGFR-TK inhibition are statistically unacceptable ($Q^2 < 0.5$).

In comparison to CoMFA, CoMSIA methodology has the advantage of exploring more fields. More statistically robust models were obtained from the CoMSIA study for all three PTKs. The best CoMSIA models for all three PTKs included hydrophobic and H-bond donor fields (CoMSIA-HD models). The best model describing PDGFR-TK inhibitory activity (CoMSIA-HD-1) has a Q^2 value of 0.807 using three components, explains 95.0% of the variance, has a low standard deviation

($s = 0.265$), and a high Fischer ratio ($F = 89.53$). The best model describing FGFR-TK inhibitory activity (CoMSIA-HD-2) has a Q^2 value of 0.703 using four components, explains 90.0% of the variance, has a low standard deviation ($s = 0.420$), and a high Fischer ratio ($F = 29.35$). Finally, the best model describing c-Src-TK inhibitory activity (CoMSIA-HD-3) has a Q^2 value of 0.853 using four components, explains 97.9% of the variance, has a low standard deviation ($s = 0.264$), and a high Fischer ratio ($F = 149.45$). All CoMSIA-HD models showed an even contribution of both hydrophobic and H-bond donor fields. For all three PTK inhibitory activities, models which only include only one field showed less reliable statistics, and the addition of other fields does not produce an improvement in the internal validation of CoMSIA-HD models. The predictions of $\log(10^6/IC_{50})$ values for the 18 SSPs in the training set using CoMSIA-HD models are shown in Table 1. The correlations between the calculated and experimental values of $\log(10^6/IC_{50})$ (from training and LOO cross-validation) are shown in Figure 3.

The contour plots of the CoMSIA hydrophobic and H-bond donor fields (stdev * coeff) are presented in Figure 4 for the modeled PTK inhibitory activities. In this figure, yellow contours represent areas where substitution of hydrophobic groups would be favorable for activity, while white contours represent regions where hydrophobicity of the substituent or placement of hydrophobic groups diminishes the PTK inhibitory activity. Besides, cyan contours represent regions where H-bond donor fields favor activity; while purple contours indicate regions where H-bond donor fields are disfavored for activity. For simplicity, the interaction between only the most active compound for each PTK and its respective contour map is shown.

Immediately afterwards, we analyze the requirements of the basic scaffold of the pyrido[2,3-*d*]pyrimidine heterocycle for increasing PTK inhibitory activities according to the contour maps. From Figure 4, it seems that the requirements for hydrophobic and H-bond donor fields are similar for PDGFR-TK and c-Src-TK inhibitory activities, meanwhile some differences are noted for FGFR-TK inhibition. All three contours showed a H-bond donor favored region (cyan isopleths) in the vicinity of the NH group at C-2. This characteristic indicates the importance of 2-NH group for providing a hydrogen

Table 3. Correlation matrix for descriptors selected by linear GA^a

	MATS1m	MATS8v	MATS6e	MATS6m	MATS1e	GATS5v	ATS1v	MATS3v	GATS1v
MATS1m	1								
MATS8v	0	1							
MATS6e	0.084	0.257	1						
MATS6m	0.508	0.025	0.122	1					
MATS1e	0.012	0.002	0.018	0.093	1				
GATS5v	0.603	0.063	0.001	0.417	0.011	1			
ATS1v	0.321	0.066	0.056	0.149	0.081	0.774	1		
MATS3v	0.226	0.008	0.005	0.012	0.127	0.199	0.259	1	
GATS1v	0.824	0.028	0.312	0.412	0.042	0.382	0.181	0.135	1

^a High correlations ($R^2 > 0.7$) are indicated in boldface.

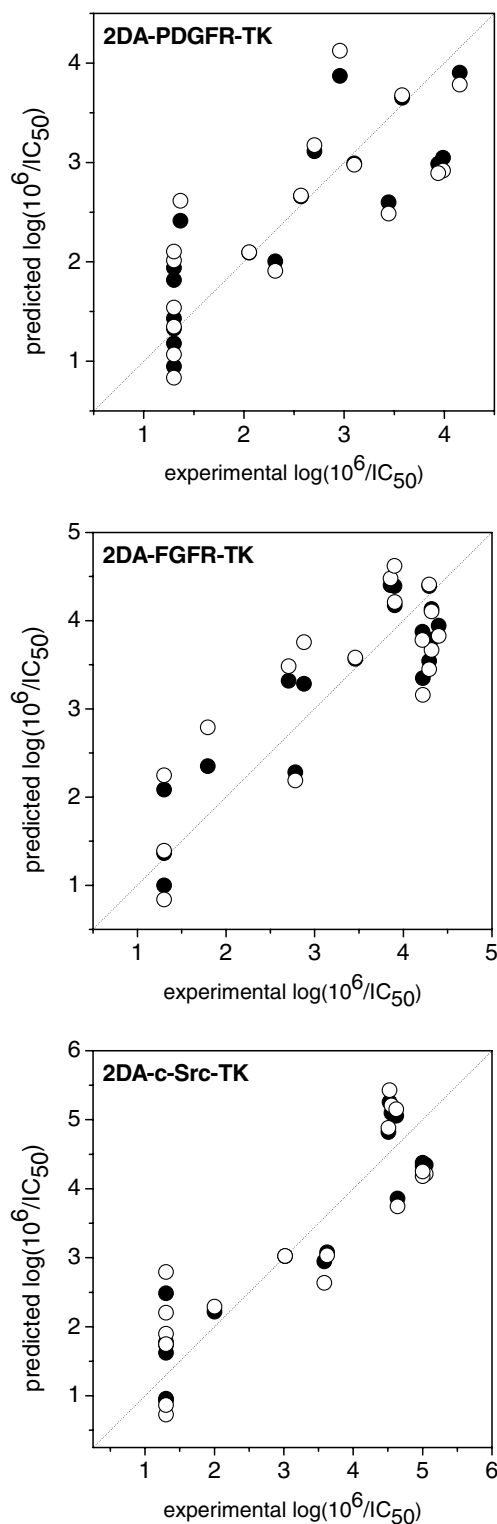


Figure 1. Plot of predicted versus experimental $\log(10^6/IC_{50})$ values for PTK inhibition by substituted pyrido[2,3-*d*]pyrimidine derivatives using 2D autocorrelation MLR models. (●) Training set predictions, (○) LOO cross-validated predictions.

bond donor; two substitutions in this N atom (like in compound **4**) or the absence of the group (compounds **2** and **3**) disfavors PTK inhibitory activities. Isoleths at the left of contour maps in Figure 4 indicate that monosubstitution of the 2-NH₂ group with large side chains improves PTK inhibitory activities. Compounds

without this requirement are less active (compounds **2**, **3**, and **15**). H-bond donors in this side chain disfavor the inhibitory activities against PDGFR-TK and c-Src-TK, and hydrophobic groups must constitute the side chain of FGFR-TK inhibitors. Furthermore, polar groups are required at the end of this side chain (white isopleths) for all three PTKs. White isopleths at the right of contour maps in Figure 4 indicate that polar groups are tolerated in the N-7 position (compounds **10** and **11**). The main difference between contour maps is shown at the isopleths surrounding the phenyl ring at C-6. Hydrophobic groups are required at positions 2 and 6 of this phenyl ring for PDGFR-TK and c-Src-TK inhibitors (yellow isopleths); however, polar groups in the positions 3 and 5 are well tolerated for FGFR-TK inhibitors (white isopleths) (compounds **17** and **18**).

Trumpp-Kallmeyer et al.¹⁷ performed 3D models of compounds **8**, **14**, and **21** docked into the model of the ATP-binding site of c-Src-TK for identifying the binding mode of the SPP inhibitors. Authors used X-ray structural information obtained from protein kinases with bound ATP and competitive inhibitors. As a result, authors found that SPPs bind in different orientation with respect to ATP. Instead, SPPs adopted the binding motif of olomoucine.¹⁹ Our CoMSIA-HD models satisfy most of the features of these models. In the proposed binding mode, a hydrogen bond is formed between the amino hydrogen at C-2 of the heterocycle and the carbonyl oxygen of Met-341 in c-Src-TK (Cys-683 and Ala-564 in PDGFR-TK and FGFR-TK). This confirms the importance of the presence of NH group at C-2. In the above-mentioned 3D models, the 2-amino substituents are located at the entrance of the ATP-binding pocket. The substituents at this position can form additional hydrophobic interactions with the peptide backbone and amino acid side chains of the extended coil stretch, as is reflected by our CoMSIA-HD models. The 3D model also showed that distal substituents off the 2-amino group extend out of the binding pocket and are exposed to solvent. In our CoMSIA-HD models it is reflected by white isopleths (hydrophilic groups enhance the activity) at the end of the chain. With respect to C-7 substituents, the 3D models showed that 7-carbonyl as well as the carbonyl function in the 7-urea substituent can form an electrostatic hydrogen bond with Lys 295 in c-Src-TK. Our CoMSIA-HD models showed that hydrophilic groups are required at this position. Finally, the 3D models showed that the 6-phenyl substituent is located in a deep pocket of moderate size adjacent to the adenine binding pocket. This pocket is made up of Ile-336 and Thr-338 on strand 5, Val-323 on strand 4, Ala-403, Asp-404, and Phe-405 on strand 8, and Glu-310 and Met-314 on the α C helix (numbers correspond to c-Src-TK). The 6-phenyl substituent is located close to Thr-338 on strand 5. Structural differences in this pocket between c-Src-TK, PDGFR-TK, and FGFR-TK may in part explain different specificity patterns observed when subtle modifications are made in the 6-phenyl substituent. Differences in this binding pocket among all three modeled PTKs are reflected in Table 5. Amino acid differences are present on strands 4, 5, and 8 in tyrosine kinases. Ile-336/678

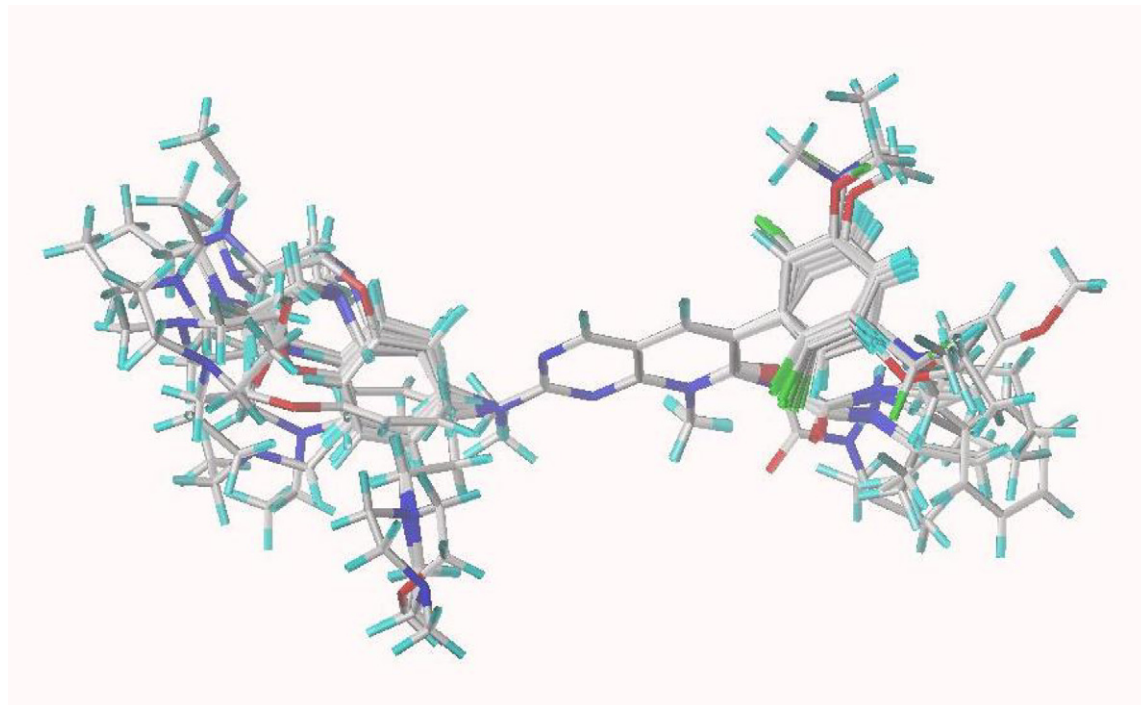


Figure 2. Atom-by-atom superposition used for 3D-QSAR analysis.

on strand 5 in c-Src-TK/PDGFR-TK is changed to Val-559 in FGFR-TK, Thr-338/680 on strand 5 of c-Src-TK/PDGFR-TK corresponds to Val-561 in FGFR-TK, Val-323/664 on strand 4 of c-Src-TK/PDGFR-TK corresponds to Ile-545 in FGFR-TK, and Ala-403/640 on strand 8 of c-Src-TK/FGFR-TK corresponds to Cys-842 in PDGFR-TK. In general, this pocket is similar for c-Src-TK and PDGFR-TK, but some differences are noted in this region for FGFR-TK. These differences change the shape of the pocket and thus might account for the selectivity patterns seen when different substituents were introduced into the 6-phenyl ring. Authors interpret that larger Ile and Thr residues which are present in PDGFR-TK and c-Src-TK might reduce the size of the pocket and thus prevent binding. However, the increase of inhibitory activity against FGFR-TK can only be noted when the 6-phenyl ring is substituted with OMe group (compounds **17** and **18**), the increase is lesser when it is substituted with OEt group (compound **16**), and there is not a significant increase when it is substituted with NMe₂ and CF₃ groups (compounds **15** and **22**). With respect to this pocket, our model CoMSIA-HD-2 identified that substituents containing hydrophilic atoms favor FGFR-TK inhibitory activity. These results indicate that changes described in Table 5 for FGFR-TK increase the availability of hydrophilic interactions at some specific regions inside the pocket.

2.3. Predictive capacity of QSAR models

High values of Q^2 are a necessary, but not a sufficient, condition for a model to possess significant predictive power. The prediction of an external test set is also an important criterion for assessing model quality.²⁰ We used 2D autocorrelation regression equations and CoMSIA-HD for each PTK inhibition to predict the inhibi-

tory activities of an external test set of inhibitors that were not used in the training process. The results of the predictions are shown in Table 1. The plots of the predicted versus experimental values and the residuals are shown in Figure 5.

This analysis reveals that all the models have a good predictive capacity. The analysis of the residuals shows that the average of 2D autocorrelation and CoMSIA-HD models leads to a more precise prediction; in this sense, it is possible to combine both models for predictive purposes. Despite the lack of interpretability of 2D autocorrelation models, such models are amenable for predicting an external set. PDGFR-TK inhibitory activities of compounds **19** and **21** are better predicted using 2DA-PDGFR-TK model, FGFR-TK inhibitory activities of compounds **19** and **22** are better predicted using 2DA-FGFR-TK model, and c-Src-TK inhibitory activities of compounds **20** and **22** are better predicted using 2DA-c-Src-TK model. On the other hand, CoMSIA-HD models are able to provide a detailed interpretation of the structure–activity relationships for the studied PTKIs as well as to predict the activity of new compounds.

3. Conclusions

Predictive QSAR models were derived for substituted pyrido[2,3-*d*]pyrimidine derivatives, which should be useful for assisting the design of selective compounds. Such models correlate well structural features with inhibitory activities against several PTKs and bring valuable information about the relevant characteristics of inhibitors.

First, 2D autocorrelation models were developed by MLR combined with GA procedure. 2D autocorrela-

Table 4. Results of the CoMFA and CoMSIA PLS analyses for PDGFR-TK, FGFR-TK, and c-Src-TK inhibitors using several different field combinations

Model	NC	R^2	S	F	Q^2	S_{CV}	Fraction				
							Steric	Electrostatic	Hydrophobic	H-bond donor	H-bond acceptor
<i>PDGFR-TK</i>											
CoMFA-S-1	6	0.975	0.214	70.3	0.614	0.833	1				
CoMFA-E-1	2	0.725	0.602	19.75	0.293	0.966		1			
CoMFA-SE-1	2	0.837	0.464	38.5	0.569	0.754	0.670	0.330			
CoMSIA-S-1	1	0.688	0.621	35.28	0.519	0.772	1				
CoMSIA-E-1	1	0.513	0.776	16.85	0.212	0.987		1			
CoMSIA-H-1	5	0.961	0.252	59.67	0.625	0.786			1		
CoMSIA-D-1	7	0.884	0.480	10.86	0.714	0.753				1	
CoMSIA-A-1	6	0.954	0.289	37.73	0.298	1.124					1
CoMSIA-SE-1	1	0.637	0.670	28.09	0.402	0.860	0.424	0.576			
CoMSIA-HD-1	3	0.950	0.265	89.53	0.807	0.536			0.426	0.574	
CoMSIA-HAD-1	4	0.974	0.199	121.74	0.769	0.593			0.464	0.416	0.120
CoMSIA-ALL-1	3	0.950	0.267	88.06	0.596	0.755	0.215	0.251	0.181	0.244	0.109
<i>FGFR-TK</i>											
CoMFA-S-2	3	0.886	0.433	36.32	0.422	0.975	1				
CoMFA-E-2	3	0.842	0.510	24.81	0.267	1.099		1			
CoMFA-SE-2	3	0.906	0.392	45.21	0.451	0.951	0.618	0.382			
CoMSIA-S-2	1	0.488	0.859	15.26	0.285	1.015	1				
CoMSIA-E-2	1	0.434	0.903	12.27	0.000	1.204		1			
CoMSIA-H-2	3	0.827	0.534	22.23	0.484	0.922			1		
CoMSIA-D-2	4	0.800	0.596	12.97	0.641	0.797				1	
CoMSIA-A-2	9	0.677	0.965	1.86	0.359	1.358					1
CoMSIA-SE-2	3	0.825	0.536	22.06	0.256	1.107	0.416	0.584			
CoMSIA-HD-2	4	0.900	0.420	29.35	0.703	0.735			0.466	0.534	
CoMSIA-HAD-2	4	0.891	0.439	26.66	0.676	0.757			0.408	0.459	0.132
CoMSIA-ALL-2	3	0.897	0.412	40.48	0.591	0.821	0.204	0.254	0.159	0.274	0.109
<i>c-Src-TK</i>											
CoMFA-S-3	2	0.865	0.620	47.99	0.724	0.887	1				
CoMFA-E-3	2	0.751	0.842	22.63	0.457	1.243		1			
CoMFA-SE-3	2	0.855	0.643	44.06	0.696	0.930	0.661	0.339			
CoMSIA-S-3	2	0.851	0.651	42.92	0.715	0.900	1				
CoMSIA-E-3	1	0.637	0.984	28.12	0.428	1.235		1			
CoMSIA-H-3	4	0.944	0.427	55.26	0.795	0.822			1		
CoMSIA-D-3	4	0.838	0.729	16.83	0.675	1.033				1	
CoMSIA-A-3	9	0.896	0.743	7.70	0.501	1.631					1
CoMSIA-SE-3	2	0.844	0.666	40.59	0.609	1.055	0.433	0.567			
CoMSIA-HD-3	4	0.979	0.264	149.45	0.853	0.705			0.549	0.451	
CoMSIA-HAD-3	4	0.971	0.311	107.30	0.836	0.733			0.481	0.432	0.087
CoMSIA-ALL-3	3	0.957	0.361	104.66	0.763	0.849	0.209	0.206	0.175	0.255	0.094

NC, is the number of components from PLS analysis; R^2 is the square of the correlation coefficient; S is the standard deviation of the regression; F is the Fischer ratio; and Q^2 and S_{CV} are the correlation coefficient and standard deviation of the leave-one-out (LOO) cross-validation, respectively. The best models are indicated in boldface.

tion space was combined with different weighting schemes, viewed as an adaptive descriptor space, containing topological information able to capture structural complexity. The 2D autocorrelation descriptors appeared to capture sufficient structural detail to yield very useful results in modeling biological properties. Our results corroborate that the employment of 2D autocorrelation descriptors is extremely useful in QSAR studies giving simple correlations between the molecular structures and biological activities.

Afterwards traditional CoMFA and CoMSIA approaches have been applied to derive the same relationships. The present study indicates that the steric and electrostatic CoMFA fields are not enough to describe fully the modeled PTK activities. When using hydrophobic and H-donor CoMSIA field, statistically meaningful models were derived. Thus, prediction of PTK inhibi-

tory activities with sufficient accuracy should be possible. Moreover, an interpretation of the CoMSIA fields makes it possible to draw conclusions concerning the most appropriate features for the analogues.

On the basis of the current results, the reported models have the potential to discover new potent and selective inhibitors and bring useful molecular information about the ligand specificity for interacting with each PTK-binding pocket.

4. Computational methods

4.1. Datasets: source and prior preparation

Inhibitions of PDGFR-TK, FGFR-TK, and c-Src-TK (IC₅₀) for 22 SPPs were taken from the literature.¹⁷

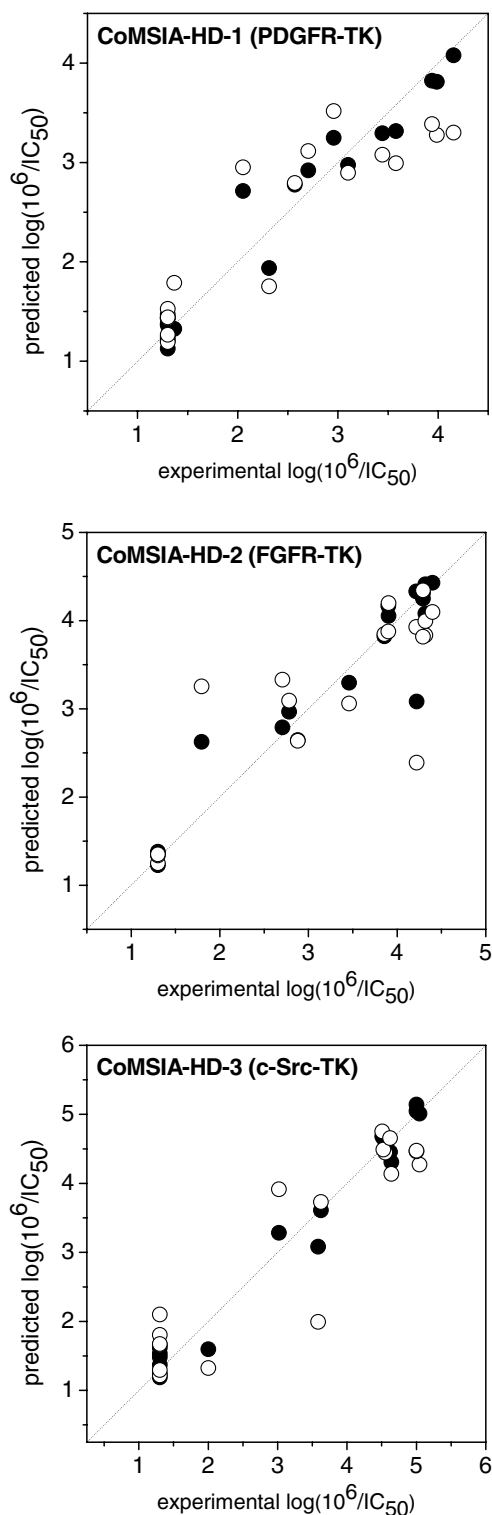


Figure 3. Plot of predicted versus experimental $\log(10^6/IC_{50})$ values for PTK inhibition by substituted pyrido[2,3-*d*]pyrimidine derivatives using CoMSIA-HD models. (●) Training set predictions, (○) LOO cross-validated predictions.

For modeling, IC_{50} activities were converted in logarithmic activities $\log(10^6/IC_{50})$, where 10^6 guarantees that logarithmic activities range between 1 and 9. The chemical structures and experimental activities ($\log(10^6/IC_{50})$) are shown in Table 1. The activity

parameters IC_{50} (μM) are measures of inhibitory activity and refer to the micromolar concentration of the PTKIs leading to 50% inhibition of the PTK. Prior to molecular descriptor calculations, 3D structures of the studied compounds were geometrically optimized using the semiempirical quantum-chemical method PM3²¹ implemented in the MOPAC 6.0 computer software.²² The dataset was divided in training and test sets. Four compounds were chosen randomly as a test set and were used for external validation for the MLR, CoMFA, and CoMSIA models. The compounds in the test sets were reserved to validate potential models. For the development of the models, the training sets included all the remaining compounds.

4.2. 2D Autocorrelation pool

Three spatial autocorrelation vectors were employed for modeling the inhibitory activities: Broto-Moreau's autocorrelation coefficients (ATS) (Eq. 4),²³ Moran's indices (MATS) (Eq. 5),²⁴ and Geary's coefficients (GATS) (Eq. 6).²⁵

$$ATS(p_k, l) = \sum_i \delta_{ij} p_{ki} p_{kj} \quad (4)$$

$$MATS(p_k, l) = \frac{N}{2L} \frac{\sum_{ij} \delta_{ij} (p_{ki} - \bar{p}_k)(p_{kj} - \bar{p}_k)}{\sum_i (p_{ki} - \bar{p}_k)} \quad (5)$$

$$GATS(p_k, l) = \frac{(N-1)}{4L} \frac{\sum_{ij} \delta_{ij} (p_{ki} - \bar{p}_k)(p_{kj} - \bar{p}_k)}{\sum_i (p_{ki} - \bar{p}_k)} \quad (6)$$

where $ATS(p_k, l)$, $MATS(p_k, l)$, and $GATS(p_k, l)$ are Broto-Moreau's autocorrelation coefficient, Moran's index, and Geary's coefficient at spatial lag l , respectively; p_{ki} and p_{kj} are the values of property k of atoms, i and j , respectively; \bar{p}_k is the average value of property k , L is the number of nonzero values in the sum, N is the number of atoms in the molecule, and $\delta(l, d_{ij})$ is a Dirac-delta function defined as

$$\delta(l, d_{ij}) = \begin{cases} 1 & \text{if } d_{ij} = l \\ 0 & \text{if } d_{ij} \neq l \end{cases} \quad (7)$$

where d_{ij} is the topological distance or spatial lag between atoms i and j .

Spatial autocorrelation measures the level of interdependence between properties, and the nature and strength of that interdependence. In a molecule, Moran's and Geary's spatial autocorrelation analysis tests whether the value of an atomic property at one atom in the molecular structure is independent of the values of the property at neighboring atoms. If dependence exists, the property is said to exhibit spatial autocorrelation. The autocorrelation vectors represent the degree of similarity between molecules. The Dragon software²⁶ was used for calculating weighted Broto-Moreau, Moran, and Geary 2D autocorrelation vectors. Four different weighting schemes have been used: atomic masses (m), atomic van der Waals volumes (v), atomic Sanderson electronegativities (e), and atom-

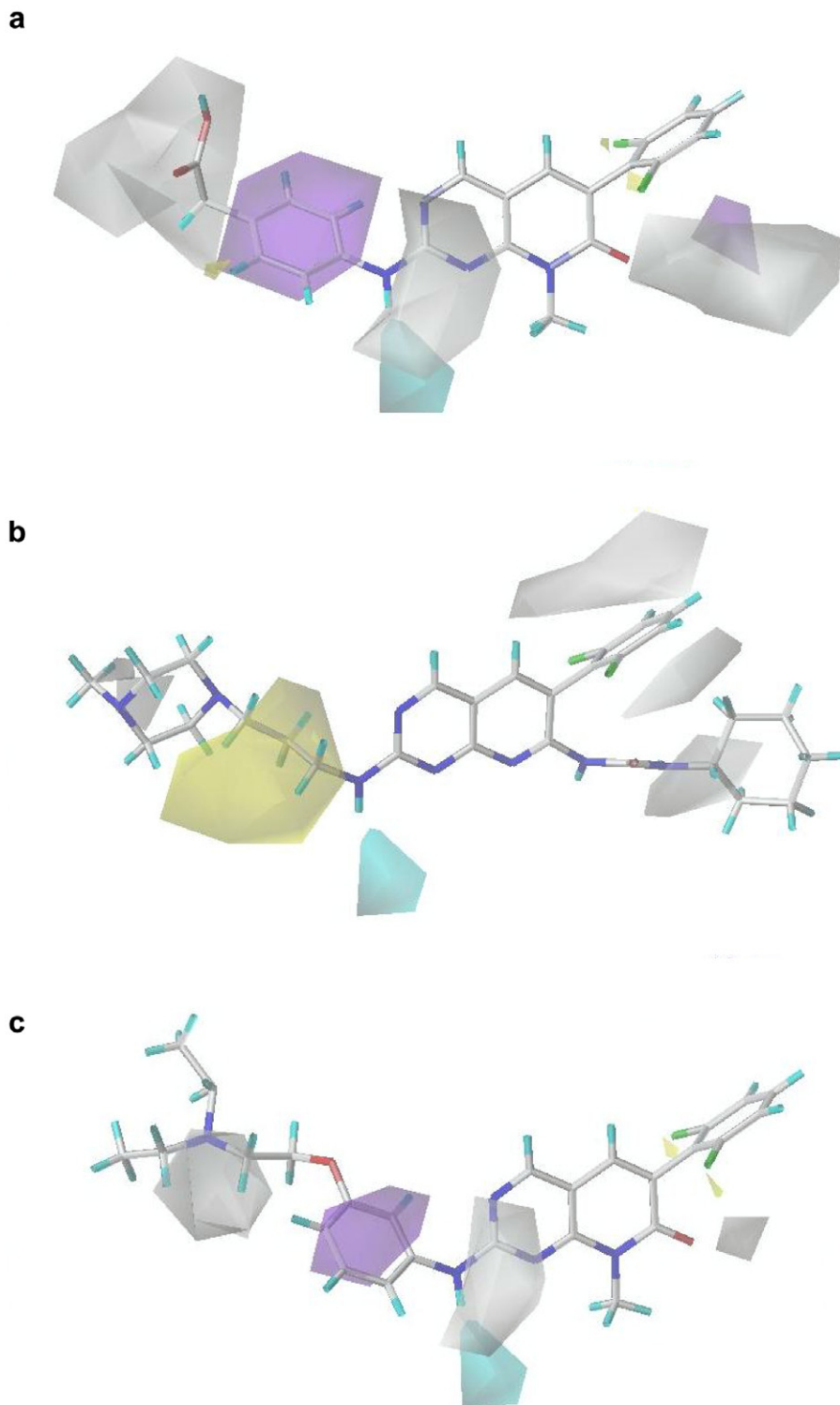


Figure 4. CoMSIA stdev * coeff contour maps for PTKIs. (a) PDGFR-TK inhibition (CoMSIA-HD-1 model), compound **8** is shown inside the field. (b) FGFR-TK inhibition (CoMSIA-HD-2 model), compound **11** is shown inside the field. (c) c-Src-TK inhibition (CoMSIA-HD-3 model), compound **7** is shown inside the field. Yellow and white isopleths indicate regions where hydrophobic and hydrophilic groups, respectively, will enhance the affinity. Cyan and purple isopleths represent areas where H-bond donors are favored and disfavored, respectively.

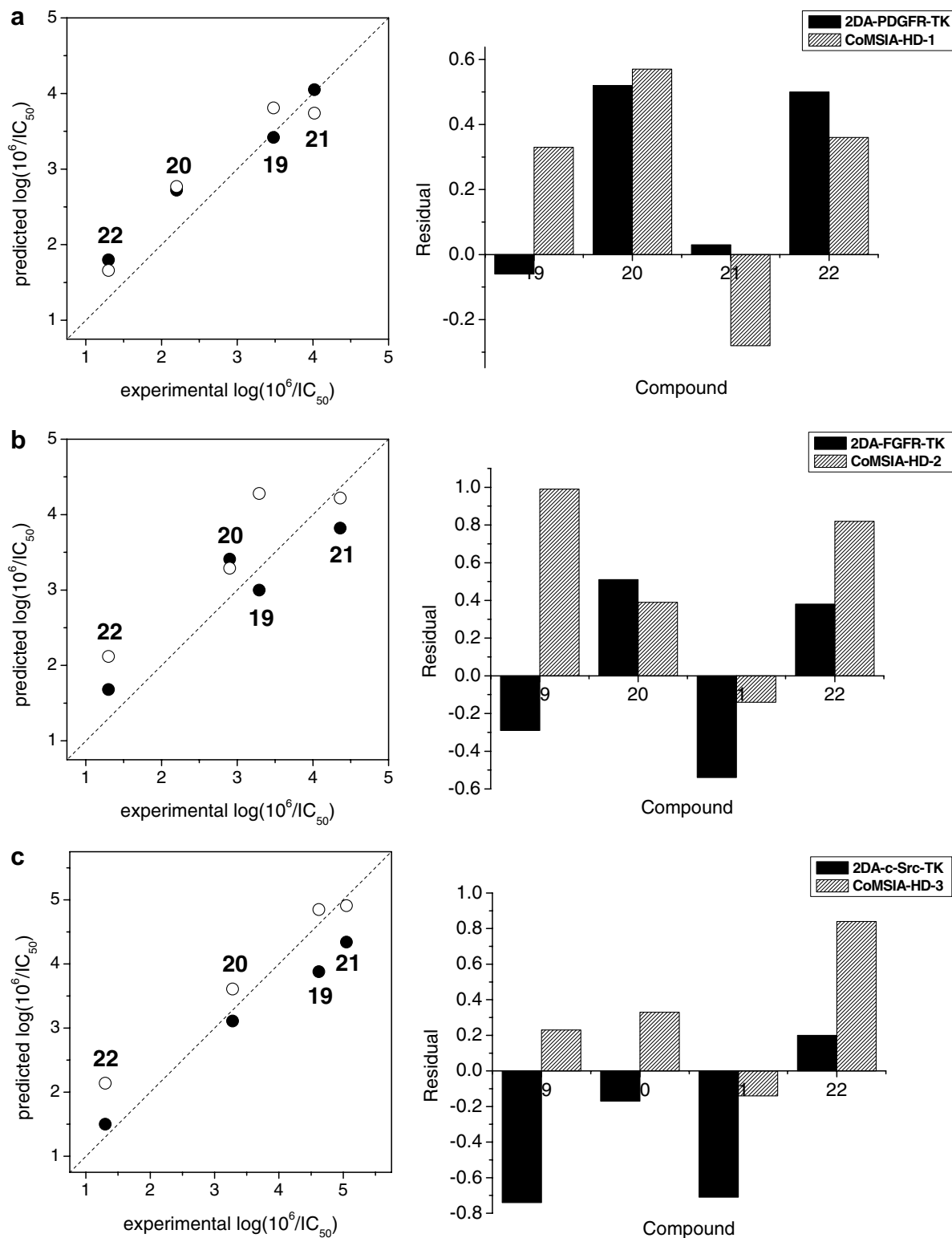
ic polarizabilities (p). Autocorrelation vectors were calculated at spatial lags l ranging from 1 up to 8. The autocorrelation descriptors are denoted by the scheme:

type of descriptor-spatial lag-weighting property; for instance, GATS6v is the Geary autocorrelation of lag 6 weighted by atomic van der Waals volumes. The

Table 5. Structural differences in the pocket which surround 6-phenyl substituent

	Strand 5	Strand 4	Strand 8
PDGFR-TK	Ile-678, Thr-680	Val-664	Cys-842
FGFR-TK	Val-559, Val-561	Ile-545	Ala-640
c-Src-TK	Ile-336, Thr-338	Val-323	Ala-403

use of 2D autocorrelation space proved advantageous in previous works.^{27–34} The definition of autocorrelation like indices is a very active field of research covering from small drugs to proteins, which has generated some recent publications including research articles and reviews.^{35,36}

**Figure 5.** Plot of predicted versus experimental $\log(10^6/IC_{50})$ values and residuals for test set compounds. (a) PDGFR-TK inhibition, (b) FGFR-TK inhibition, (c) c-Src-TK inhibition. (●) 2DA and (○) CoMSIA-HD models.

A data matrix was generated with the spatial autocorrelation vectors calculated for each compound. Afterward, dimensionality reduction methods were employed for selecting the most relevant vector components for building MLR models. The total number of computed descriptors was 96. Descriptors with constant values were discarded. For the remaining descriptors, pairwise correlation analysis was performed in order to reduce, in a first step, the colinearity and correlation between descriptors. The procedure consists of the elimination of the descriptor with lower variance from each pair of descriptors with the modulus of the pair correlation coefficients higher than a predefined value ($R^2_{\max} = 0.95$). Afterward, the number of remained descriptors was 39.

4.3. MLR modeling procedure

Since many molecular descriptors were available for QSAR analysis and only a reduced subset of them is statistically significant in terms of correlation with biological activities, deriving an optimal QSAR model through variable selection needs to be addressed. Following the Occam's Razor,³⁷ we selected just the variables that contain the information that is necessary for the modeling but nothing more. In this sense, GA searches have been carried out in order to build the MLR models. The quality of each model was proven by the square multiple correlation coefficient (R^2) and the standard deviation (S). The models with R^2 -value above 0.8 were selected and they were tested in cross-validation experiments.

GA search was carried out exploring MLR models. The mean square error of data fitting was tried as the individual fitness function. An initial population of 50 individuals is randomly extracted from the data matrix in the first generation. The succeeding generations were generated by crossover and single-point mutation operators, while the two best scoring individuals were automatically retained as members for the next round of evolution. The GA search ends when 90% of the generations showed the same target fitness score. Linear GA was programmed within the MATLAB environment using the genetic algorithm toolbox.³⁸ The best models were selected according to R value ($R > 0.8$) and the results of cross-validation experiments (higher Q^2).

4.4. CoMFA and CoMSIA models

CoMFA and CoMSIA models were obtained using the Sybyl 7.2 software of Tripos.³⁹ All the molecules were aligned by an atom-by-atom least-square fit. We used the 5-vinylpyrimidine fragment of the active compound **8** in its optimized conformation as a template. For the 3D-QSAR calculations previous dataset splitting was kept. The molecules of the training set were placed in a rectangular grid extended beyond 4 Å in each direction from the coordinates of each molecule. The interaction energies between a probe atom (a sp^3 hybridized carbon atom with +1 charge) and all compounds were computed at the surrounding points, using a volume-dependent lattice with 2.0 Å grid spacing. Then, standard Sybyl parameters were used for a partial least squares (PLS) analysis. The number of components in the PLS

models was optimized by using Q^2 value, obtained from the LOO cross-validation procedure, with the SAMPLS⁴⁰ sampling method. The number of components was increased until additional components did not increase Q^2 by at least 5% per added component. The CoMFA models were generated by using steric and electrostatic probes with standard 30 kcal/mol cut-offs. In the CoMSIA analyses, similarity is expressed in terms of steric occupancy, electrostatic interactions, local hydrophobicity, and H-bond donor and acceptor properties, using a 0.3 attenuation factor.

4.5. Analysis of the quality of the models

The quality of the fit of the training set of a specific model was measured by its R^2 . However, a most important measure is the prediction quality. An internal LOO cross-validation process was carried out by estimating R^2 of LOO cross-validation (Q^2) and standard deviation (S_{CV}). A data point was removed (left-out) from the training set, and the model was refitted; the predicted value for that point is then compared to its actual value. This is repeated until each datum has been omitted once; the sum of squares of these deletion residuals can then be used to calculate Q^2 . In addition, the predictive power of the model was also measured by an external validation process that consists in predicting the activity of unknown compounds forming the test set. In this case, the residuals of predictions were evaluated. Such criteria have been formulated as the requirements for a QSAR model to have highly predictive power.

Acknowledgment

This work was supported by 'Programa Bicentenario de Ciencia y Tecnología', ACT/24 (JC and FDGN) and Innova Chile (FDI, CORFO), Project 06CN12PAT-51 (JC and FDGN).

References and notes

1. Biscardi, J. S.; Tice, D. A.; Parsons, S. J. *Adv. Cancer Res.* **1999**, *76*, 61.
2. Parsons, J. T.; Martin, K. H.; Slack, J. K.; Taylor, J. M.; Weed, S. A. *Oncogene* **2000**, *19*, 5606.
3. Levitzki, A. *Acc. Chem. Res.* **2003**, *36*, 462.
4. Supuran, C. T.; Scozzafava, A. *Expert Opin. Ther. Pat.* **2004**, *14*, 35.
5. Wakeling, A. E. *Curr. Opin. Pharmacol.* **2002**, *2*, 382.
6. Druker, B. J. *Semin. Hematol.* **2003**, *40*, 50.
7. Slamon, D. J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; Baselga, J.; Norton, L. N. *Engl. J. Med.* **2001**, *344*, 783.
8. Ferrara, N.; Hillan, K. J.; Gerber, H.-P.; Novotny, W. *Nat. Rev. Drug Disc.* **2004**, *3*, 391.
9. González-Díaz, H.; Saiz-Urra, L.; Molina, R.; Santana, L.; Uriarte, E. *J. Proteome Res.* **2007**, *6*, 904.
10. González-Díaz, H.; Saiz-Urra, L.; Molina, R.; González-Díaz, Y.; Sánchez-González, A. *J. Comput. Chem.* **2007**, *28*, 1042.
11. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: New York, 1993.

12. Novič, M.; Nikolovska-Coleska, Z.; Solmajer, T. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 990.
13. Kurup, A.; Garg, R.; Hansch, C. *Chem. Rev.* **2001**, 101, 2573.
14. Shen, Q.; Lu, Q.-Z.; Jiang, J.-H.; Shen, G.-L.; Yu, R.-Q. *Eur. J. Pharm. Sci.* **2003**, 20, 63.
15. Guha, R.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 2179.
16. Chen, G.; Luo, X.; Zhu, W.; Luo, C.; Liu, H.; Puah, C. M.; Chena, K.; Jiang, H. *Bioorg. Med. Chem.* **2004**, 12, 2409.
17. Trumpp-Kallmeyer, S.; Rubin, R. J.; Humblet, C.; Hamby, J.; Showalter, H. D. S. *J. Med. Chem.* **1998**, 41, 1752.
18. Cronin, M. T. D.; Schultz, T. W. *J. Mol. Struct. (Theochem)* **2003**, 622, 39.
19. Schulze-Gahmen, U.; Brandsen, J.; Jones, H. D.; Morgan, D. O.; Meijer, L.; Vesely, J.; Kim, S.-H. *Proteins* **1995**, 22, 378.
20. Golbraikh, A.; Tropsha, A. *J. Mol. Graph. Model.* **2002**, 20, 269.
21. Stewart, J. J. P. *J. Comput. Chem.* **1989**, 10, 210.
22. MOPAC version 6.0. U.S. Air Force Academy: Colorado Springs, CO.
23. Moreau, G.; Broto, P. *Nouv. J. Chim.* **1980**, 4, 359.
24. Moran, P. A. P. *Biometrika* **1950**, 37, 17.
25. Geary, R. F. *Incorporated Statistician* **1954**, 5, 115.
26. Todeschini, R.; Consonni, V.; Pavan, M. DRAGON, version 2.1. Talet SRL: Milan, Italy.
27. Fernández, M.; Tundidor-Camba, A.; Caballero, J. *Mol. Simul.* **2005**, 31, 575.
28. Fernández, M.; Caballero, J. *Bioorg. Med. Chem.* **2006**, 14, 280.
29. Caballero, J.; Garriga, M.; Fernández, M. *Bioorg. Med. Chem.* **2006**, 14, 3330.
30. Fernández, M.; Caballero, J. *J. Mol. Graph. Model.* **2006**, 25, 410.
31. González, M. P.; Caballero, J.; Helguera, A. M.; Garriga, M.; González, G.; Fernández, M. *B. Math. Biol.* **2006**, 68, 735.
32. Fernández, M.; Caballero, J.; Tundidor-Camba, A. *Bioorg. Med. Chem.* **2006**, 14, 4137.
33. Caballero, J.; Tundidor-Camba, A.; Fernández, M. *QSAR Comb. Sci.* **2007**, 26, 27.
34. Fernández, M.; Caballero, J. *Bioorg. Med. Chem.* **2007**, 15, 6298.
35. Agüero-Chapin, G.; González-Díaz, H.; Molina, R.; Varona-Santos, J.; Uriarte, E.; González-Díaz, Y. *FEBS Lett.* **2006**, 580, 723.
36. González-Díaz, H.; Vilar, S.; Santana, L.; Uriarte, E. *Curr. Top. Med. Chem.* **2007**, 7, 1025.
37. Hawkins, D. M. *J. Chem. Inf. Comput. Sci.* **2004**, 44, 1.
38. MATLAB, version 7.0. The Mathworks Inc.: Natick, MA, <http://www.mathworks.com>.
39. SYBYL, version 7.2; Tripos Inc.: 1699 South Hanley Road, St. Louis, MO 63144, USA.
40. Bush, B. L.; Nachbar, R. B., Jr. *J. Comput.-Aided Mol. Des.* **1993**, 7, 587.