Chemometrics and Intelligent Laboratory Systems 134 (2014) 47–52

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

# Conformation-independent QSAR on c-Src tyrosine kinase inhibitors

Nieves C. Comelli [a,*], Erlinda V. Ortiz [b], Magdalena Kolacz [c], Alla P. Toropova [d], Andrey A. Toropov [d], Pablo R. Duchowicz [c], Eduardo A. Castro [c]

[a] Facultad de Ciencias Agrarias, Universidad Nacional de Catamarca, Av. Belgrano y Maestro Quiroga No. 50, 4700 Catamarca, Argentina
[b] Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca, Av. Maximio Victoria 55, 4700 Catamarca, Argentina
[c] Instituto de Investigaciones Fisicoquímicas Teóricas y Aplicadas INIFTA (UNLP, CCT La Plata-CONICET), Diag. 113 y 64, C.C. 16, Sucursal 4, 1900 La Plata, Argentina
[d] Istituto di Ricerche Farmacologiche Mario Negri, 20156, Via La Masa 19, Milano Italy

ARTICLE INFO

*Article history:*
Received 2 October 2013
Received in revised form 24 January 2014
Accepted 2 March 2014
Available online 11 March 2014

*Keywords:*
QSAR theory
Multivariable linear regression analysis
c-Src tyrosine kinase
Structural descriptors
Pyrrolo-pyrimidine derivatives

ABSTRACT

The main idea of this work was to find predictive quantitative structure–activity relationships (QSAR) for a wide set of c-Src tyrosine kinase inhibitors, by means of resorting to a conformation-independent representation of the chemical structure. In this way, our attempt was to avoid the availability of X-ray crystallographic structural information of the target. Therefore, in a set composed of 80 pyrrolo-pyrimidine derivatives, 1179 theoretical descriptors were simultaneously analyzed through linear regression models obtained with the replacement method variable subset selection technique. Alternatively, the flexible (activity dependent) descriptor approach was also applied in this study. The models were validated and tested through the use of an external test set of compounds, the leave-group-out cross validation method, Y-randomization and applicability domain analysis. Our results were compared with previously published ones based on docking analysis and 3D-QSAR. The obtained conformation-independent approach was in good agreement with experimental observations.



## 1. Introduction

The c-Src enzyme is the most studied non-receptor class of tyrosine kinase (TK), that along with the Yes, Lck, Fyn, Lyn, Fgr, Hck, Blk and Yrk members constitutes the Src (sarcoma) family[1,2]. As participant of signal transduction chains in cells, c-Src kinase catalyzes the transfer of phosphate from adenosine triphosphate (ATP) to specific tyrosine residues of the target substrate [3–5]. The overexpression, deregulation or mutation of c-Src associated with alterations in its activity have been observed in various cancer types, such as colon, breast, and pancreatic types [6,7]. The oncogenic potential of an increased Src activity involves the control of cell proliferation and the regulation of cytoskeletal-linked events, such as migration, spreading, and invasion [8–11]. Considerable evidences also suggest that c-Src kinase inhibition may enhance the anti-tumor efficacy of hormonal and cytotoxic agents in preclinical models [12–14].

Several c-Src TK inhibitors have been identified during past years, which include the following heterogeneous classes: dihydro-pyrimidoquinolinones [15], pyrazolo-pyrimidines [16], pyrrolo-pyrimidines [17–20], 4-anilinoquinazolines [21], and others [22]. In particular, pyrrolo-pyrimidines interfere with the c-Src role on osteoclastic bone resorption and metastatic bone disease [23]. These inhibitors are ATP-competitive, in contrast to other compounds which display a more complex behavior (as substrate competitive or bisubstrate competitive against either the substrate or ATP) [22].

The increasing efforts for identifying and developing new selective and potent small-molecule inhibitors of c-Src have focused on the integration of information generated by different ways [14,24]: research of designing, synthesizing and testing more potent biomolecules has been generally assisted by molecular modeling techniques. Structure–activity relationships (SAR) and quantitative structure–activity relationships (QSAR) help to predict the biological activity of new structures and may reveal useful information on structural modification at several substitutional positions of c-Src-binding molecules [24–29]. The QSAR theory is based on the main hypothesis that the biological activity of a chemical compound is a final result of its molecular structure [30]. This assumption does not offer specific details on the usually complex mechanism/path of action involved although; however, it is possible to get insight on the underlying mechanism by means of the QSAR-based predicted activities.

It is known that 3D-QSAR is a broad term encompassing all QSAR methods which correlate macroscopic target properties with molecular descriptors derived from the spatial three-dimensional molecular structure [31,32]. However, the application of these methods suffers from serious drawbacks, some of which are: i) the lack of accurate X-ray crystal structural information of the drug–target complex, ii) the binding mode of the drug is unknown or highly promiscuous, iii) the drugs are not analogs and alignment hypothesis is difficult to establish [33,34]. In this context, the conformation-independent 1D and 2D QSAR methods emerge as an alternative approach for developing

* Corresponding author. Tel.: +54 383 4435955; fax: +54 383 4430504.
*E-mail address:* ncomelli75@gmail.com (N.C. Comelli).

http://dx.doi.org/10.1016/j.chemolab.2014.03.003
0169-7439/

models which include the commonly important features related to dataset and the specific contributions of each derivative type.

In a recent study, Tintori et al. [26] have reported molecular docking simulations and a 3D-QSAR analysis on a wide set of pyrrolo-pyrimidine inhibitors, using the 3D structure of c-Src as template for structure-based alignment and by applying the GRID/GOLPE approach. The identified model has a good statistical quality and is capable of predicting the activity in an external test set of compounds; it is also applied over a set of pyrazolo-pyrimidine derivatives. The gathered information provides useful insight into the rational design of more efficient compounds and explicit the interaction types that should be established by c-Src TK inhibitors. In addition, it has been investigated the pharmacokinetic features of all inhibitors under study by projecting the compounds on predefined Volsurf models, allowing for some important considerations and suggestions to direct the development of new inhibitors.

In this study we establish QSAR models on the same dataset of pyrrolo-pyrimidine inhibitors considered in [26], with the purpose of applying an approach that does not consider the conformational representation of the chemical structure [35,36]. The exclusion of 3D structural aspects also avoids problems associated with ambiguities that result from an incorrect computational geometry optimization due to the existence of compounds in various conformational states. Such kind of problems may also lead to the loss of predictive capability of the QSAR when applied for the prediction of an external test set of compounds. Therefore, it is expected that the established model would be useful for predicting the c-Src TK inhibition whenever the molecular descriptors involved in it reflect an appropriate parallelism between the observed activity and the chosen molecular characterization. There is no similar QSAR study on the specific dataset of pyrrolo-pyrimidine inhibitors.

## 2. Materials and methods

### 2.1. Experimental dataset

The chemical domain analyzed involved 80 pyrrolo-pyrimidines (please see Table 1S of Supplementary Tables) [26]. The experimental c-Src TK inhibitory activity measured as $pIC_{50} = -\log_{10} IC_{50}$ ranged from 5.0 to 9.5. The activity values were obtained from the liquid-phase tyrosine phosphorylation assay, using chicken c-Src and the synthetic substrate poly-Glu-Tyr [20].

### 2.2. Molecular descriptor calculation

The compounds were first drawn with HyperChem for Windows [37], and then the.hin files were converted into.sdf by using the Open Babel 2.3.1 software [38]. We computed 931 conformation-independent molecular descriptors using E-Dragon [39]. This well-known descriptor database was comprised thirteen descriptor families: Constitutional, Topological, Walk and Path Count, Connectivity Index, Information Index, Edge Adjacency Index, Topological Charge Index, Burden Eigenvalues, Eigenvalue-Based Index, 2D-Autocorrelation, Functional Group Count, Atom-Centred Fragment, and Molecular Property [40].

In addition, we obtained 248 Transferable Atom Equivalent (TAE) descriptors using Recon 5.5 [41]. This sort of descriptors was not provided by E-Dragon, while the robustness of Recon has previously been demonstrated elsewhere [42,43]. Recon is an algorithm for the reconstruction of molecular charge densities and charge density-based electronic properties of molecules, using atomic charge density fragments precomputed from ab initio wavefunctions. The method is based on the Quantum Theory of Atoms in Molecules [44]. A library of atomic charge density fragments has been assembled in a form that allows for the rapid retrieval of the fragments, followed by rapid molecular assembly. The.sdf molecular format was employed as input for the generation of 248 Transferable Atom Equivalent (TAE) descriptors developed by Breneman et al. [45]. In this way, the total number of calculated molecular descriptors was 1179 variables.

### 2.3. Model development

#### 2.3.1. Molecular descriptor selection in MLR

In recent years theoretical and experimental researchers focused their attention on finding the most efficient tools for selecting molecular descriptors in QSAR studies. A great number of feature selection methods to search for the best descriptors from a pool of variables is available, and the replacement method (RM) [46,47], employed here, was successfully applied elsewhere [48–52]. In brief, the RM has been an efficient optimization tool which generates multivariable linear regression (MLR) based models on the training set by searching in a set having $D$ descriptors for an optimal subset having $d << D$ ones with smallest training set standard deviation ($S_{train}$) or smallest root mean square deviation ($RMSD_{train}$). The quality of the results achieved with this technique approaches that obtained by performing an exact (combinatorial) full search of molecular descriptors although, of course, it requires much less computational work. However, in some cases, the RM can get trapped in a local minimum of $S$. Although such local minima provide acceptable models, an improvement of the method has been developed in the enhanced replacement method (ERM) [53,54]. We used Matlab 7.12 in all our calculations [55].

#### 2.3.2. The CORAL method

CORAL version 1.5 [56] is a freeware for Windows, where each compound is represented with the SMILES notation (Simplified Molecular Input Line Entry System), calculated here by converting the.hin files into.smi with Open Babel (see Table 2S). A CORAL model is a one-variable linear correlation between a given biological activity and a flexible (activity dependent) descriptor (*DCW*). This flexible descriptor was calculated in this work by choosing a graph representation for chemical structure on the training set, for which three options were available: hydrogen-suppressed graph, hydrogen-filled graph and graph of atomic orbitals. The *DCW* is a linear combination of special coefficients, the so-called correlation weights (*CW*). The *CW* are calculated for structural attribute values on the training set, such as the values taken by the hierarchy of Morgan's extended connectivity indices of $k^{th}$ order for a given vertex $i$ ($^{k}EC_i$, $k = 0$–3).

The Monte Carlo (MC) simulation method was used as a way for obtaining the *CW*, by searching for the highest correlation coefficient ($R$) between the biological activity and the *DCW* descriptor in a number of different probes (3). The *DCW* values depend upon the threshold value ($T$) and the number of epochs ($N_{epochs}$) used, which are positive integer parameters of the Monte Carlo method that should be correctly specified in order to calculate the *DCW* values. The $T$ defines rare (noise) SMILES attributes that do not contribute to the predicted activity, so that all SMILES attributes that take place in less than $T$ SMILES notations of the training set are classified as rare instead of as active. More specific details of the CORAL algorithm are available in the recent literature [57–60].

#### 2.3.3. Model validation

The most reliable model validation consists on using an external test set of molecules. The same training and test set partition from ref. [26] was used in present analysis for comparison purposes. The training set involved 65 pyrrolo-pyrimidines compounds (see Table 1S). These compounds were selected following the usual guidelines: compounds belonging to the training set should be representative of the molecular diversity of all the compounds under study and uniformly span over the whole range of activity. The remaining compounds (15 pyrrolo-pyrimidines) were used as external test.

We practiced the cross-validation technique of leave-one-out (loo) and leave-more-out (ln%o, with n% being the percentile of molecules removed from the training set). The statistical parameters $R_{ln\%o}$ and $S_{ln\%o}$ (correlation coefficient and standard deviation of leave-more-

out) measured the stability of the QSAR upon inclusion/exclusion of molecules. The number of cases for random data removal analyzed was 50,000. According to the specialized literature, a necessary but not sufficient condition for the loo explained variance ($R^2_{loo}$) is to be greater than 0.5 for a validated model [61].

The Y-randomization procedure [62] was applied in order to verify that the model is robust. This technique consists on scrambling the experimental property values in such a way that they do not correspond to the respective compounds. After analyzing 10,000 cases of Y-randomization, the standard deviation obtained ($S^{rand}$) has to be a poorer value than the one found by considering the true calibration ($S$).

*2.3.4. Applicability domain analysis*

The applicability domain (AD) for the QSAR model was also explored, as not even a predictive model is expected to reliably predict the modeled activity for the whole universe of molecules. The AD is a theoretically defined area that depends on the descriptors and the experimental activity [63]. Only the molecules falling within this AD were not considered model extrapolations. One possible way to characterize the AD is based on the leverage approach [64], which allows to verify whether a new compound can be considered as interpolated (with reduced uncertainty, reliable prediction) or extrapolated outside the domain (unreliable prediction). Each compound $i$ has a calculated leverage value ($h_i$) and there exists a warning leverage value ($h^*$); Table 2S includes the definitions for $h_i$ and $h^*$. When $h_i > h^*$ for a test set compound, then a warning should be given: it means that the prediction is the result of substantial extrapolation of the model and could not be treated as reliable.

*2.3.5. Degree of contribution of selected descriptors*

In order to find out the relative importance of the *j-th* descriptor in the linear model, we standardized its regression coefficient ($b^s_j$, see Table 2S). The larger the absolute value of $b^s_j$, the greater the importance of such descriptor [65].

## 3. Results and discussion

Table 4S summarizes the best MLR models found having 1–10 variables and the meaning of each involved descriptor is supplied by Table 5S. The results suggest that the best model is given by 5 descriptors obtained with E-Dragon, which has an acceptable predictive power on the test set. In E-Dragon based descriptors, the $RMSD_{train}$ parameter continues improving beyond such number of 5 variables, but $RMSD_{test}$ becomes poorer. We follow the common practice of keeping the model's size as small as possible, in order to avoid any possible fortuitous correlation. It is also noted from Table 4S that, as $d$ increases, Recon descriptors have a lesser fit to the training set when compared with E-Dragon performance, and that the combination of E-Dragon and Recon descriptors tend to overfit the training set. According to this, we choose the following structure–activity relationship:

$$pIC_{50} = 6.579\,IC3 - 5.864IC5 + 0.712\,MATS7m - 229.473\,JGI7 + 1.339nArOH + 6.685 \quad (1)$$

$N_{train} = 65$, $d = 5$, $N_{train}/d = 13$, $R^2_{train} = 0.71$, $S_{train} = 0.61$, $F = 27.38$, $R^2_{ij\ max} = 0.86$, $o(2.5\ S) = 1$, $R^2_{loo} = 0.64$, $S_{loo} = 0.67$, $R^2_{l20\%o} = 0.46$, $S_{l20\ \%\ o} = 0.83$, $S^{rand} = 0.90$, $N_{test} = 15$, $R^2_{test} = 0.58$, $S_{test} = 0.87$.

Here, $F$ is the Fisher parameter, $R_{ij\ \max}$ denotes the maximum correlation coefficient between descriptor pairs, $o(2.5\ S)$ indicates the number of outlier compounds having a residual (difference between experimental and calculated activity) greater than 2.5 times $S_{train}$ and lower than three times $S_{train}$.

In Eq. (1), the percentage of explained variances are $R^2_{train} = 71\%$ and $R^2_{test} = 58\%$. This model has a predictive performance on the external test set that compares fairly well with that found by Tintori et al. [26], with $R^2_{train} = 94\ \%$ and $R^2_{test} = 71\ \%$. In addition, the root mean square deviations (*RMSD*) for the present study are: $RMSD_{train} = 0.58$ and $RMSD_{test} = 0.68$, while the ones obtained by using 3D-QSAR are $RMSD_{train} = 0.26$ and $RMSD_{test} = 0.55$. A two-tailed significance test, *t*-test for dependent groups or paired *t*-test [66], reveals that there is no *statistically significant difference* between *two* sets of predicted response of the external test set and their statistics (mean of the difference response, $\overline{X}_D = 0.087$; standard deviation of the difference responses, $S_D = 0.201$; two-tailed critical and calculated $t$ values at $p = 0.05$ with 14 degrees of freedom, $t_{crit.}(14) = 2.145$ and $t_{calc.}(14) = 0.430$; then $t_{calc.}(14) < t_{crit.}(14)$).

On the order hand, the model given by Eq. (1) approves the internal validation process of cross-validation through the exclusion of one molecule at a time and also by excluding 20% of the observations (13 molecules). We apply Y-randomization, demonstrating that $S_{train} < S^{rand}$ and thus a valid structure–activity relationship is achieved. We have checked that Eq. (1) accomplishes with the external validation criteria recommended in [61] to assure predictive capability:

- $1 - R^2_0/R^2_{test} < 0.1$ (0.004) or $1 - R'^2_0/R^2_{test} < 0.1$ (0.36) and,
- $0.85 \le k \le 1.15$ (0.97) and $0.85 \le k' \le 1.15$ (1.03)
- $R^2_m > 0.5$ (0.55)

where $R^2_0$ and $R'^2_0$ are the coefficients of determination for regressions through the origin in the test set of the observed versus predicted activities and the predicted and observed activities respectively; $k$ and $k'$ are the slopes of regression lines through the origin and $R^2_m$ is the modified squared correlation coefficient. These parameters were calculated as shown in Table 3S.

Fig. 1 plots the predicted $pIC_{50}$ activity as a function of the experimental values for the training and test sets (numerical data provided in Table 6S), showing that there exists a tendency for the points to have a straight line trend. The dispersion plot of residuals (i.e. residuals as a function of predicted activities) in Fig. 1S reveals that residuals tend to obey a random pattern around the zero line, suggesting that the assumption of the MLR technique is fulfilled.

One compound with a value of residual greater than 2.5 times $S_{train}$ is present in the training set, compound **8**. After checking that their experimental $pIC_{50}$ values and structures are correct, we assume that this abnormal behavior may be purely attributed to the structural diversity of the dataset. It has been also identified previously [26] that **8**, as well as **9**–**11**, **34**, **75**–**77** and **79**, have a different binding mode within the ATP binding pocket of the active c-Src TK domain. This could be a possible explanation of this outlier behavior.

Regarding the descriptors appearing in Eq. (1) there are:

- two Information indices: *IC*3, Information Content index (neighbourhood symmetry of 3-order) and *IC*5, Information Content index (neighbourhood symmetry of 5-order). These descriptors, obtained from the elements of Graph Theory, describe the connectivity and branching in a molecule.

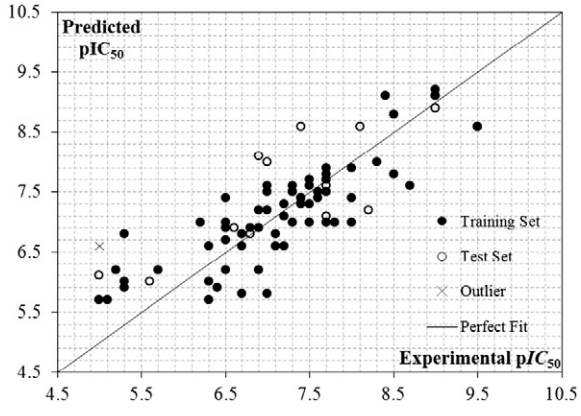


**Fig. 1.** Predicted and experimental $pIC_{50}$ values for training and test set with Eq. (1).

- two 2D-Autocorrelation descriptors: *MATS7m*, Moran autocorrelation of lag 7 weighted by mass and *JGI7*, Mean topological charge index of order 7. These indices, derived from weighted matrices in terms of atomic masses and electronegativities, address the topology of the structure or parts of it codifying chemical information about atom type and bond multiplicity.
- one Functional Group Count: *nArOH*. This descriptor gives local chemical information about the number of aromatic hydroxyls in a molecule.

The contribution degree for each descriptor ($b_j^s$) reveals that *IC*3 and *IC*5 contribute more to the predictions: *IC*3(1.8), *IC*5 (1.3), *MATS7m*(0.1), *JGI7* (0.7), and *nArOH*(0.6). Analyzing the more relevant descriptors, as these take positive numerical values, it is concluded that the sign of each regression coefficient in Eq. (1) affects the predicted activity value. Thus, higher values for *IC*3, *nArOH* and lower values for *IC*5 and *JGI7* would lead to more potent compounds (higher predicted $pIC_{50}$ activities).

The descriptors in Eq. (1) embody chemical features that map all molecules in the training and test sets as heterocyclic compounds with polarizable hydrophobic substituents and ring aromatic with hydrogen-bond acceptor. These structural features also were identified as guidelines for optimization strategies and in SAR studies of related compounds [17–20]. Thus, the information obtained from the Eq. (1) QSAR model may provide the tools for predicting the affinity modifying of related compounds and for guiding future structural synthesizing new potent c-Src inhibitors.

On the other hand, the maximum correlation coefficient between the *IC*3 and *IC*5 descriptors is $R_{ij\ \max} = 0.93$ (see correlation matrix in Table 7S). We understand that, although there is a high correlation between such variables, these are not collinear, and each of them includes different aspects of the molecular structure [67] that succeed in combining with the remaining variables of Eq. (1). The numerical values taken by the five descriptors are included in Table 8S.

An analysis of the AD of Eq. (1) reveals that the 15 compounds included in the test set belong to the AD of the model. The leverage values are provided in Table 9S.

We compute flexible descriptors as an alternative modeling methodology in the current QSAR study. After trying different combinations of ${}^kEC_j$ indices ($k = 0$–3), the best correlation weights produced by the Monte Carlo simulation (listed in Table 10S) are obtained from a hydrogen-filled graph representation, leading to the following model:

$$pIC_{50} = 0.262\,DCW - 1.377 \qquad (2)$$

$N_{train} = 65$, $d = 1$, $N_{train}/d = 65$, $R^2_{train} = 0.70$, $S_{train} = 0.58$, $F = 149.49$,$o(2.5\ S) = 0$, $R^2_{loo} = 0.68$, $S_{loo} = 0.60$, $R^2_{l20\%o} = 0.55$, $S_{l20\ \%\ o} = 0.72$, $S^{rand} = 0.96$,$N_{test} = 15$, $R^2_{test} = 0.81$, $S_{test} = 0.54$.

The parameters used during model building are $T = 6$ ($T$ ranges from 0 to 7) and $N_{epochs} = 30$. Table 11S includes an example for calculating *DCW* in **1**. The *DCW* descriptor of Eq. (2) considers ${}^3EC_j$ as local graph invariant, and from a total number of 118 structural attributes (Table 10S), 56 of them contribute to *DCW* calculation in this set of pyrrolo-pyrimidines. Furthermore, higher positive *CW* values tend to predict higher $pIC_{50}$ activities, while higher negative *CW* values tend to predict less potent compounds. This CORAL QSAR model has $RMSD_{train} = 0.57$ and $RMSD_{test} = 0.50$, with a predictive capability on the test set comparable to the published result [26]. Eq. (2) has no outliers exceeding 2.5 times the *S* value (see Figs. 2 and 2S). All the 15 test set compounds belong to the AD of this model, having $h^* = 0.0923$. Eq. (2) also satisfies the external validation conditions reported in [61]:

- $1 - R^2_0/R^2_{test} < 0.1$ (2.41 × 10$^{-5}$) and $1 - R'^2_0/R^2_{test} < 0.1$ (0.05) and,
- $0.85 \le k \le 1.15$ (0.96) and $0.85 \le k' \le 1.15$ (1.04)
- $R^2_m > 0.5$ (0.81)

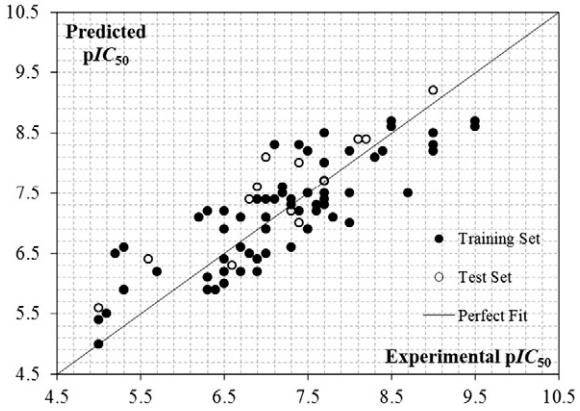


**Fig. 2.** Predicted and experimental $pIC_{50}$ values for training and test set with Eq. (2).

## 4. Conclusions

In this work, we investigate the relationship between the chemical structure of a set of 80 pyrrolo-pyrimidine derivatives and their capability to inhibit the c-Src TK activity. The application of linear models for selecting the most relevant structural parameters results in the combination *IC*3, *IC*5, *MATS7m*, *JGI7*, and *nArOH* Dragon descriptors. The importance of these variables in modeling the studied property is validated and tested ($R^2_{loo} = 0.64$, $R^2_{test} = 0.58$ and $RMSD_{test} = 0.68$). The acceptable predictive power on the test set is compared to previously published and more sophisticated 3D-QSAR results.

In addition, we report an alternative linear model with base in the flexible descriptor definition. The statistical performance of the CORAL QSAR model is $R^2_{loo} = 0.64$, $R^2_{test} = 0.68$ and $RMSD_{test} = 0.50$.

The reported models have acceptable predictive power on the test set. Our results were compared to previously published and more sophisticated 3D-QSAR results.

Currently, we are working on the application of these and others QSAR models based on constitutional or topological approximations as virtual screening tools to discover novel and potential inhibitors against c-Src tyrosine kinase from databases and/or virtual chemical libraries. New results will be published soon elsewhere.

## Acknowledgments

We thank the financial support provided by the National Research Council of Argentina (CONICET) PIP11220100100151 project and the Ministerio de Ciencia, Tecnología e Innovación Productiva for the electronic library facilities. N.C.C. acknowledges a fellowship from CONICET and Facultad de Ciencias Agrarias, Universidad Nacional de Catamarca. MK is grateful to a fellowship from the International Association for the Exchange of Students for Technical Experience (IAESTE).

## Appendix A. Supplementary data.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.chemolab.2014.03.003.

## References

[1] D.H. Drewry, P. Bamborough, K. Schneider, G.K. Smith, The kinome and its impact on medicinal chemistry, in: R.A. Ward, F. Goldberg (Eds.), Kinase Drug Discovery, Royal Society of Chemistry, W1J 0BA, Burlington House, Piccadilly, London, 2011, pp. 1–53.
[2] T. Hunter, Tyrosine phosphorylation: thirty years and counting, Curr. Opin. Cell Biol. 21 (2009) 140–146.

[3] T.J. Boggon, M.J. Eck, Structure and regulation of Src family kinases, Oncogene 23 (2004) 7918–7927.
[4] S.J. Parsons, J.T. Parsons, Src family kinases, key regulators of signal transduction, Oncogene 23 (2004) 7906–7909.
[5] B. Sen, F.M. Johnson, Regulation of SRC family kinases in human cancers, J. Signal Transd. 2011 (2011) 865819–865832.
[6] A.Y. Tsygankov, S.K. Shore, Src: regulation, role in human carcinogenesis and pharmacological inhibitors, Curr. Pharm. Des. 10 (2004) 1745–1756.
[7] J.M. Summy, G.E. Gallick, Src family kinases in tumor progression and metastasis, Cancer Metastasis Rev. 22 (2003) 337–358.
[8] R.B. Irby, T.J. Yeatman, Increased Src activity disrupts cadherin/catenin-mediated homotypic adhesion in human colon cancer and transformed rodent cells, Cancer Res. 62 (2002) 2669–2674.
[9] N. Rahimi, W. Hung, E. Tremblay, R. Saulnier, B. Elliott, c-Src kinase activity is required for hepatocyte growth factor-induced motility and anchorage-independent growth of mammary carcinoma cells, J. Biol. Chem. 273 (1998) 33714–33721.
[10] R.J. Jones, E. Avizienyte, A.W. Wyke, D.W. Owens, V.B. Brunton, M.C. Frame, Elevated c-Src is linked to altered cell-matrix adhesion rather than proliferation in KM12C human colorectal cancer cells, Br. J. Cancer 87 (2002) 1128–1135.
[11] V.G. Brunton, E. Avizienyte, V.J. Fincham, B. Serrels, C.A. Metcalf, T.K. Sawyer, Frame identification of Src-specific phosphorylation site on focal adhesion kinase: dissection of the role of Src SH2 and catalytic functions and their consequences for tumor cell behavior, Cancer Res. 65 (2005) 1335–1342.
[12] M.S. Duxbury, H. Ito, M.J. Zinner, S.W. Ashley, E.E. Whang, Inhibition of SRC tyrosine kinase impairs inherent and acquired gemcitabine resistance in human pancreatic adenocarcinoma cells, Clin. Cancer Res. 10 (2004) 2307–2318.
[13] A. Levitzk, A. Gazit, Tyrosine kinase inhibition: an approach to drug development, Science 267 (1995) 1782–1788.
[14] F.A. Al-Obeidi, K.S. Lam, Development of inhibitors for protein tyrosine kinases, Oncogene 19 (2000) 5690–5701.
[15] R.L. Dow, B.M. Bechle, T.T. Chou, C. Goddard, E.R. Larson, Selective inhibition of the tyrosine kinase pp60src by analogs of 5,10-dihydropyrimido[4,5-b]quinolin-4(1H)-one, Bioorg. Med. Chem. Lett. 5 (1995) 1007–1010.
[16] F. Carraro, A. Naldini, A. Pucci, G.A. Locatelli, G. Maga, S. Schenone, O. Bruno, A. Ranise, F. Bondavalli, C. Brullo, P. Fossa, G. Menozzi, L. Mosti, M. Modugno, C. Tintori, F. Manetti, M. Botta, Pyrazolo[3,4-d]pyrimidines as potent antiproliferative and proapoptotic agents toward A431 and 8701-BC cells in culture via inhibition of c-Src phosphorylation, J. Med. Chem. 49 (2006) 1549–1561.
[17] E. Altmann, L. Widler, M. Missbach, N7-substituted-5-aryl-pyrrolo[2,3-d]pyrimidines represent a versatile class of potent inhibitors of the tyrosine kinase c-Src, Mini-Rev. Med. Chem. 2 (2002) 201–208.
[18] M. Missbach, E. Altmann, L. Widler, M. Susa, E. Buchdunger, H. Mett, T. Meyer, J. Green, Substituted 5,7-diphenyl-pyrrolo[2,3d]pyrimidines: potent inhibitors of the tyrosine kinase c-Src, Bioorg. Med. Chem. Lett. 10 (2000) 945–949.
[19] L. Widler, J. Green, M. Missbach, M. Susa, E. Altmann, 7-Alkyl- and 7-cycloalkyl-5-aryl-pyrrolo[2,3-d]pyrimidines—potent inhibitors of the tyrosine kinase c-Src, Bioorg. Med. Chem. Lett. 11 (2001) 849–852.
[20] E. Altmann, M. Missbach, J. Green, M. Susa, H.A. Wagenknecht, L. Widler, 7-Pyrrolidinyl- and 7-piperidinyl-5-aryl-pyrrolo[2,3-d]pyrimidines—potent inhibitors of the tyrosine kinase c-Src, Bioorg. Med. Chem. Lett. 11 (2001) 853–856.
[21] M.R. Myers, N.N. Setzer, A.P. Spada, A.L. Zulli, C.J. Hsu, A. Zilberstein, S.E. Johnson, L.E. Hook, M.V. Jacoski, The preparation and sar of 4-(anilino), 4-(phenoxy), and 4-(thiophenoxy)-quinazolines: inhibitors of p56lck and EGF-R tyrosine kinase activity, Bioorg. Med. Chem. Lett. 7 (1997) 417–420.
[22] S. Schenone, S. Zanoli, C. Brullo, E. Crespan, G. Maga, Current advances in the development of anticancer drugs targeting tyrosine kinases of the Src family, Curr. Drug Ther. 3 (2008) 158–176.
[23] B.F. Boyce, L. Xing, Z. Yao, T. Yamashita, W.C. Shakespeare, Y. Wang, C.A. Metcalf, R. Sundaramoorthi, D.C. Dalgarno, J.D. Iuliucci, T.K. Sawyer, SRC inhibitors in metastatic bone disease, Clin. Cancer Res. 12 (2006) 6291s–6295s.
[24] M.H. Bohari, H.K. Srivastava, G.N. Sastry, Analogue-based approaches in anti-cancer compound modelling: the relevance of QSAR models, Org. Med. Chem. Lett. 1 (2011) 3.
[25] S.K. Bairy, B.V. Suneel Kumar, J.U. Bhalla, A.B. Pramod, M. Ravikumar, Three-dimensional quantitative structure-activity relationship studies on c-Src inhibitors based on different docking methods, Chem. Biol. Drug Des. 73 (2009) 416–427.
[26] C. Tintori, M. Magnani, S. Schenone, M. Botta, Docking, 3D-QSAR studies and in silico ADME prediction on c-Src tyrosine kinase inhibitors, Eur. J. Med. Chem. 44 (2009) 990–1000.
[27] R. Cao, N. Mi, H. Zhang, 3D-QSAR study of c-Src kinase inhibitors based on docking, J. Mol. Model. 16 (2010) 361–375.
[28] H. Cao, H. Zhang, X. Zheng, D. Gao, 3D QSAR studies on a series of potent and high selective inhibitors for three kinases of RTK family, J. Mol. Graph. Model. 26 (2007) 236–245.
[29] M. Fernandez, S. Ahmad, A. Sarai, Proteochemometric recognition of stable kinase inhibition complexes using topological autocorrelation and support vector machines, J. Chem. Inf. Model. 50 (2010) 1179–1188.
[30] M.T.D. Cronin, Quantitative structure–activity relationships (QSARs). Applications and methodology, in: T. Puzyn, J. Leszczynski, M.T.D. Cronin (Eds.), Recent Advances in QSAR Studies: Methods and Applications, Springer Science & Business Media, New York, 2010.
[31] T.I. Oprea, 3D QSAR modeling in drug design, in: P. Bultinck, H.D. Winter, W. Langenaeker, J.P. Tollenaere (Eds.), Computational Medicinal Chemistry for Drug Discovery, Marcel Dekker, Inc., New York, USA, 2004, pp. 571–616.
[32] J. Verma, V.M. Khedkar, E.C. Coutinho, 3D-QSAR in drug design—a review, Curr. Top. Med. Chem. 10 (2010) 95–115.
[33] T. Puzyn, J. Leszczynski, M.T.D. Cronin, Recent advances in QSAR studies: methods and applications, Challenges and Advances in Computational Chemistry and Physics, Springer Science&Business Media B.V., Netherlands, 2010.
[34] J. Subramanian, S. Sharma, C. B-Rao, A novel computational analysis of ligand-induced conformational changes in the ATP binding sites of cyclin dependent kinases, J. Med. Chem. 49 (2006) 5434–5441.
[35] A. Talevi, C.L. Bellera, M.D. Ianni, P.R. Duchowicz, L.E. Bruno-Blanch, E.A. Castro, An integrated drug development approach applying topological descriptors, Curr. Comput. Aided Drug Des. 8 (2012) 172–181.
[36] P.R. Duchowicz, N.C. Comelli, E.V. Ortiz, E.A. Castro, QSAR study for carcinogenicity in a large set of organic compounds, Curr. Drug Saf. 7 (2012) 282–288.
[37] HyperChem, version 7.5, Hypercube, Inc., 2002. (http://www.hyper.com).
[38] N.M. O'Boyle, M. Banck, C.A. James, C. Morley, T. Vandermeersch, G.R. Hutchison, Open Babel: an open chemical toolbox, J. Cheminf. 3 (2011) 1–14.
[39] E-Dragon, version 1.0,VCCLAB, Virtual Computational Chemistry Laboratory, 2005. (http://michem.disat.unimib.it/chm).
[40] R. Todeschini, V. Consonni, Molecular descriptors for chemoinformatics, Weinheim, 2009.
[41] Recon, version 5.5, Rensselaer Polytechnic Institute, New York, USA, 2002. (http://reccr.chem.rpi.edu/Software/RECON/recondoc/).
[42] B.K. Lavine, C.E. Davidson, C. Breneman, W. Katt, Electronic Van der Waals surface property descriptors and genetic algorithms for developing structure-activity correlations in olfactory databases, J. Chem. Inf. Comput. Sci. 43 (2003) 1890–1905.
[43] A. Worachartcheewan, C. Nantasenamat, T. Naenna, C. Isarankura-Na-Ayudhya, V. Prachayasittikul, Modeling the activity of furin inhibitors using artificial neural network, Eur. J. Med. Chem. 44 (2009) 1664–1673.
[44] R.F.W. Bader, Atoms in Molecules—A Quantum Theory, Clarendon Press, U.K., 1990.
[45] C.M. Breneman, L.W. Weber, Transferable atom equivalents. Assembling accurate electrostatic potential fields for large molecules from ab initio and PROAIMS results on model systems, in: G.A. Jeffrey, J.F. Piniella (Eds.), The Application of Charge Density Research to Chemistry and Drug Design, Plenum, New York, 1991.
[46] P.R. Duchowicz, E.A. Castro, F.M. Fernández, Alternative algorithm for the search of an optimal set of descriptors in QSAR-QSPR studies, MATCH Commun. Math. Comput. Chem. 55 (2006) 179–192.
[47] P.R. Duchowicz, E.A. Castro, F.M. Fernández, M.P. González, A new search algorithm of QSPR/QSAR theories: normal boiling points of some organic molecules, Chem. Phys. Lett. 412 (2005) 376–380.
[48] P.R. Duchowicz, A. Talevi, L.E. Bruno-Blanch, E.A. Castro, New QSPR study for the prediction of aqueous solubility of drug-like compounds, Bioorg. Med. Chem. Lett. 16 (2008) 7944–7955.
[49] M. Goodarzi, P.R. Duchowicz, C.H. Wu, F.M. Fernández, E.A. Castro, New hybrid genetic based Support Vector Regression as QSAR approach for analyzing flavonoids-GABA(A) complexes, J. Chem. Inf. Model. 49 (2009) 1475–1485.
[50] A.B. Pomilio, M.A. Giraudo, P.R. Duchowicz, E.A. Castro, QSPR analyses for aminograms in food: citrus juices and concentrates, Food Chem. 123 (2010) 917–927.
[51] A. Talevi, M. Goodarzi, E.V. Ortiz, P.R. Duchowicz, C.L. Bellera, G. Pesce, E.A. Castro, L. E. Bruno-Blanch, Prediction of drug intestinal absorption by new linear and non-linear QSPR, Eur. J. Med. Chem. 46 (2011) 218–228.
[52] G. Pasquale, G.P. Romanelli, J.C. Autino, J. García, E.V. Ortiz, P.R. Duchowicz, Quantitative structure-activity relationships on chalcone derivatives: mosquito larvicidal studies, J. Agric. Food Chem. 60 (2012) 692–697.
[53] A.G. Mercader, P.R. Duchowicz, F.M. Fernandez, E.A. Castro, Replacement method and enhanced replacement method versus the genetic algorithm approach for the selection of molecular descriptors in QSPR/QSAR theories, J. Chem. Inf. Model. 50 (2010) 1542–1548.
[54] A.G. Mercader, P.R. Duchowicz, F.M. Fernandez, E.A. Castro, Advances in the replacement and enhanced replacement method in QSAR and QSPR theories, J. Chem. Inf. Model. 51 (2011) 1575–1581.
[55] Matlab, version 7.12.0.635, The MathWorks Inc., Masachussetts, USA, 2011. (http://www.mathworks.com).
[56] A.A. Toropov, E. Benfenati, Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: using rare SMILES attributes to define the applicability domain, Bioorg. Med. Chem. 26 (2008) 4801–4809.
[57] A.A. Toropov, E. Benfenati, SMILES in QSPR/QSAR modeling: results and perspectives, Curr. Drug Discov. Technol. 4 (2007) 77–116.
[58] A.A. Toropov, E. Benfenati, Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: using rare SMILES attributes to define the applicability domain, Bioorg. Med. Chem. 26 (2008) 4801–4809.
[59] A.A. Toropov, A.P. Toropova, E. Benfenati, Simplified molecular input line entry system-based optimal descriptors: quantitative structure-activity relationship modeling mutagenicity of nitrated polycyclic aromatic hydrocarbons, Chem. Biol. Drug Des. 73 (2009) 515–525.
[60] A.A. Toropov, A.P. Toropova, E. Benfenati, D. Leszczynska, J. Leszczynski, InChI-based optimal descriptors: QSAR analysis of fullerene[C60]-based HIV-1 PR inhibitors by correlation balance, Eur. J. Med. Chem. 45 (2010) 1387–1394.
[61] K. Roy, On some aspects of validation of predictive quantitative structure-activity relationship models, Expert. Opin. Drug Discov. 2 (2007) 1567–1577.
[62] S. Wold, L. Eriksson, Statistical validation of QSAR results, in: H. van de Waterbeemd (Ed.), Chemometrics Methods in Molecular Design, Weinheim VCH, 1995, pp. 309–318.
[63] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR Comb. Sci. 26 (2007) 694–701.

[64] L. Eriksson, J. Jaworska, A.P. Worth, M.T. Cronin, R.M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs, Environ. Health Perspect. 111 (2003) 1361–1375.
[65] N.R. Draper, H. Smith, Applied Regression Analysis, John Wiley and Sons, New York, 1981.
[66] J.A. Rice, Mathematical Statistics and Data Analysis, Duxbury Press, Belmont, California, 1995.
[67] P.R. Duchowicz, F.M. Fernández, E.A. Castro, Orthogonalization methods in QSPR—QSAR studies, QSPR-QSAR Studies on Desired Properties for Drug Design, 2010, pp. 189–203