MIA-QSAR modelling of anti-HIV-1 activities of some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners

Matheus P. Freitas*

Received 18th November 2005, Accepted 26th January 2006
First published as an Advance Article on the web 6th February 2006
DOI: 10.1039/b516396j

A QSAR method based on MIA (multivariate image analysis) descriptors is applied to a series of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners, compounds with anti-HIV-1 activity. Two models were built in order to appraise the modelling capability when different drawing programs are used to create the set of molecules. Both models showed good predictive ability, with cross-validated Q^2 of 0.712 and 0.624, and Q^2 for an external validation set of 0.823 and 0.747. An ADME evaluation, by calculating the topological polar surface area (TPSA) and parameters derived from the rule of five, was also performed to proposed compounds in order to suggest absorption profiles for potential new drugs.

Introduction

Computer-assisted approaches to the design of drugs, known as QSAR and QSPR (quantitative structure-activity/property relationships), have been used for many years in order to avoid intuitive and empirical ways to achieve an active or leader compound. The virtual procedure may be capable of predicting potential new drugs with a high confidence, since in addition to QSAR techniques, tools for prediction of ADME-Tox parameters (absorption, distribution, metabolism, excretion and toxicity) are available and becoming more numerous and popular nowadays.

The most refined methods of drug design are based on 3D approaches, since information about non-covalent effects, such as steric, electrostatic and hydrogen bonding interactions, may be found by using these methodologies. CoMFA,[1] CoMSIA[2] and GRID[3] methods are some of the important means utilized to model compounds of specific biological interests. However, 3D methods require, in general, knowledge about main conformation(s) and alignment of ligands in order to generate suitable descriptors to represent the interactions. The conformational screening is often an exhaustive procedure, especially to molecules with a large number of degrees of freedom and size, and the 3D alignment step is sometimes too complicated to be performed, since more than one conformation of a molecule may have low energy enough to be chosen as a possible ligand, and a correct center should be used to superimpose all the conformationally unconstrained molecules.

In the 1960's, Free and Wilson[4] developed a mathematical model to correlate structures with their respective activities. Free and Wilson[4] were pioneers in this field, but even today it remains difficult to obtain results which are comparable to those achieved from 3D and 4D approaches (4D formalism has already been invoked in some studies[5]). Promising attempts to simplify and offer advantages over 3D methods have been emerging, such as one and two-dimensional approaches based on strings and binary descriptors,[6,7] in addition to the methodologies based on the use of classical descriptors.[8]

The recently developed MIA-QSAR (multivariate image analysis applied to QSAR) method[9] has provided a rapid analysis and results as reliable as the most sophisticated methodologies available today, and is inexpensive and facile to handle. It was applied here to a series of anti-HIV-1 compounds in order to demonstrate one application and the predictive power of the MIA-QSAR method, as well as consolidate this approach as a real possibility in the course of developing simple QSAR methods. Furthermore, parameters derived from the "rule of five" were obtained for proposed compounds with expected good activities, by using the Molinspiration program,[10] in order to obtain some ADME perspectives for these compounds.

The acquired immuno deficiency syndrome (AIDS) epidemic has claimed more than three million lives in 2004, and an estimated five million people have acquired the human immunodeficiency virus (HIV) in 2004, bringing to nearly 40 million the number of people globally living with the virus.[11] Given these alarming numbers and the recent interest in the development of novel and potent inhibitors for the treatment of the HIV-1 infection, some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners (Fig. 1), anti-HIV-1 compounds, were utilized as models in this study, whose experimental data of anti-HIV-1 activity (assayed in MT-4 cell line) were taken from a recent work of Leonard and Roy.[12]

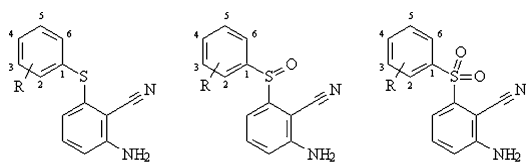

Fig. 1 2-Amino-6-arylsulfonylbenzonitriles (**35–64**) and their thio (**1–19**) and sulfinyl (**20–34**) congeners.

Departamento de Química, Universidade Federal de Lavras, Caixa Postal 37, 37200-000, Lavras, MG, Brazil. E-mail: matheus@ufla.br; Fax: +55 35 3829 1271; Tel: +55 35 3829 1271

Computational methods

The MIA-QSAR method

In the MIA-QSAR method,[9] descriptors, which are parameters required to correlate objects (in our case, the molecules) with the respective dependent variables (the biological activities), are pixels of 2D images, which are the proper structures of the molecule set drawn with aid of any appropriate program. The drawing programs used were ChemWindow[13] and the ChemSketch module of ACDLabs program.[14] Each compound of Fig. 1 (64 molecules) was drawn by using both programs, in order to demonstrate the predictive ability of models built when using different drawing software, and consequently different font type and size, and chemical group representation (for instance, Me or CH_3).

Each 2D structure was saved in Paint Brush as bitmaps, with resolution of 81 × 81 points per inch. The bitmaps were cut to a size of 250 × 205 pixels and the molecules were manually and systematically fixed in a given coordinate by a common point among them, as illustrated in Fig. 2, just as a 2D alignment of the figures. The molecules used in the model should have some similarity in molecular structure so that one can proceed with calibration, in this case using a congeneric series. Each image was unfolded to a 1 × 51250 row and then the 64 images were grouped to form a matrix of dimension 64 × 51250. Columns with zero variance were deleted, reducing the size of matrices to 64 × 1518 (when using the set of molecules drawn with the ChemWindow program) and to 64 × 1705 (when using the set of molecules drawn with the ChemSketch program), in order to minimize memory and optimize the computing cost. Finally, calibration and test set matrices were created by dividing the former 64 × 1518 matrix into a 48 × 1518 and another 16 × 1518 matrix, respectively, analogously to what occurred for the matrix with dimension 64 × 1705, which generated a training set matrix with dimension 48 × 1705 and a test set matrix of 16 × 1705.

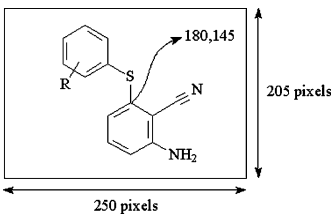

Fig. 2 Coordinates for the 2D alignment of figures and window size.

PLS regression

The key to this analysis is the reorganization (matricizing) of the original three-way array. Unfolding is done so that pixels become a single row and, thus, an image that was originally I by J pixels for K compounds is reshaped to form a two-way array that is I × J by K. An **X**-matrix is then built where each row contains the variables (the pixels) describing each molecule, and is subsequently decomposed into a score vector $\boldsymbol{s}_1$ and a weight vector $\boldsymbol{w}_1$. The score vector is determined to have the property of maximum covariance with the dependent variable y. The score vectors then replace the original variables as regressors. In the present analysis, the only data pre-processing applied to the data set was column mean-centering.

The quality of the calibrations was quantified with R^2 (training set), the external predictions with Q^2 (test set) and the cross-validated predictions with Q^2_{CV}, the squared correlation coefficients of the linear regression of experimental *versus* predicted pIC_{50} plots. The F-statistic and t-test values were also obtained to evaluate the model quality. Root mean square errors were calculated according to the following eqn (1):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(y_p - y_r)^2}{n}} \quad (1)$$

where y_p is the predicted value, y_r is the reference value and n is the number of samples.

RMSEC is the error of calibration, whilst the error of prediction (test set) and cross-validation were named RMSEP and RMSECV, respectively.

The leave-one-out cross-validation was performed with the NIPALS algorithm[15] and the calculations were carried out using the Matlab[16] platform.

Results and discussion

The set of molecules utilized in this study is a series of compounds with activity against the HIV-1 virus, some 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners, as recently reported by Chan *et al.*[17] The experimental activity data (pIC_{50}) described for the 64 compounds of Tables 1 and 2 may also be found in reference 12. The test set was chosen in such a way that compounds with low, moderate and high activities are distributed in roughly equal proportions.

Two models were taken into account in the present study: model 1, consisting of a set of molecules drawn by using the ChemWindow program,[13] and model 2, in which the molecules were drawn by using the ChemSketch program.[14] Although the set of molecules drawn by utilizing the two drawing programs is exactly the same, the font type and size used to represent atoms and groups, as well as group designation, such as $–CH_3$ or –Me to illustrate a methyl group, differ from one program to the other. Thus, calculations were carried out using both models to evaluate the difference of the descriptors on the calibration and prediction data. Since molecules are systematically drawn and aligned, if the 1518 descriptors of model 1 were correlated with dependent variables exactly as the 1705 descriptors of model 2, the difference between the calibration data of the two models is negligible, ideally null.

Figs. 3 and 4 show the good predictive abilities obtained when using both models. While model 1 predicted activity values for an external validation set (test set) with a squared correlation coefficient of 0.823 when compared with the experimental data, as can easily be seen in Table 3, model 2 presented a Q^2 of 0.747, using three latent variables (the minimum RMSEP was estimated for three latent variables in both models). In the leave-one-out cross-validation, Q^2_{CV} of 0.712 and 0.624 for models 1 and 2, respectively, also indicates the remarkable predictive power of the method, though model 1 was a bit more parsimonious (the minimum cumulative predictive residual error sum of squares was reached at three latent variables, against four for model 2). Models 1 and 2 presented reasonably different modelling capabilities, with the former showing to be a little better in this case, indicating that

Table 1 Experimental and predicted pIC_{50} values[a] for the 2-amino-6-arylsulfonylbenzonitriles (**35–64**) and their thio (**1–19**) and sulfinyl (**20–34**) congeners, using model 1

Cpd	R	Obs.	Fitted	Pred.	CV	Cpd	R	Obs.	Fitted	Pred.	CV
1	H	1.836	1.911		1.848	**33**	3-Cl, 5-Me	3.495	3.799		3.640
2	2-OMe	2.367	2.287		2.061	**34**	3-OMe, 5-CF_3	2.684	2.568		2.605
3	3-OMe	2.222	2.028		1.978	**35**	H	2.699	2.615		2.618
4	2-Me	1.796	1.656		1.594	**36**	2-OMe	3.222	2.992		2.812
5	3-Me	2.215	2.445		2.426	**37***	3-OMe	3.046		2.546	2.763
6	4-Me	0.939	0.913		1.099	**38**	4-OMe	1.602	1.984		2.575
7	2-Cl	2.387	2.009		1.767	**39**	2-Me	2.638	2.360		2.357
8*	3-Cl	2.131		2.513	2.568	**40***	3-Me	3.398		2.963	3.134
9	2-Br	1.523	1.769		1.726	**41**	4-Me	2.022	1.984		1.315
10*	3-Br	2.292		2.695	2.755	**42**	2-Cl	2.387	2.360		2.667
11	3-F	2.009	1.948		1.963	**43**	3-Cl	3.229	1.214		3.297
12	3-CN	2.762	2.005		1.814	**44**	4-Cl	2.523	2.714		1.849
13*	4-CN	1.359		0.679	0.957	**45**	2-Br	2.301	2.474		2.504
14*	3-CF_3	1.893		1.692	1.836	**46**	3-Br	3.268	3.586		3.501
15	3-NH_2	1.502	1.706		1.939	**47**	4-Br	1.699	1.825		2.118
16	3,5-Me_2	3.367	3.591		3.521	**48**	2-F	2.523	2.550		2.544
17*	3-Cl, 5-Me	2.754		3.660	3.781	**49***	3-F	2.523		2.363	2.544
18	3-OMe, 5-Me	2.699	3.175		3.280	**50***	2-CN	2.268		2.563	2.705
19	3-OMe, 5-CF_3	2.292	2.615		2.724	**51**	3-CN	2.620	2.710		2.818
20*	2-OMe	2.319		2.036	2.095	**52**	4-CN	1.097	1.167		1.645
21	3-OMe	1.796	1.964		2.096	**53**	3-CF_3	2.456	2.583		2.669
22	2-Me	1.032	1.591		1.829	**54**	2,5-Cl_2	3.523	3.209		3.026
23	3-Me	1.534	2.380		2.613	**55***	3,5-Cl_2	4.155		3.635	3.588
24*	4-Me	1.310		0.679	1.021	**56**	3,5-Me_2	5.000	4.296		4.121
25	2-Br	1.407	1.705		1.775	**57***	3-Br, 5-Me	4.699		4.546	4.508
26	3-Br	4.097	2.817		2.327	**58**	3-Cl, 5-Me	4.523	4.551		4.364
27*	4-Br	1.694		1.290	1.538	**59**	3-OMe, 5-Me	4.301	3.880		3.834
28	2-CN	2.409	1.981		1.751	**60**	3-OMe, 5-CF_3	4.046	3.320		3.213
29	3-CN	1.848	1.941		2.043	**61**	3-OH, 5-Me	3.367	3.844		4.023
30	3-CF_3	1.398	1.814		1.984	**62**	3-OCH_2CH_3, 5-Me	4.222	4.217		4.072
31*	3,5-Me_2	3.469		3.357	3.545	**63***	3-$O(CH_2)_2CH_3$, 5-Me	4.222		3.487	3.707
32	2,5-Cl_2	2.007	2.457		2.483	**64**	3-$O(CH_2)_3CH_3$, 5-Me	3.222	3.628		4.065

[a] Fitted values for the training set, predicted values (Pred.) for the test set, and leave-one-out cross-validation values (CV).

the way in which the substituents are represented has a sensible influence on the modelling. Considering that the number and "shape" of the descriptors differ substantially between the two models, the deviation observed in the results was expected.

A classical QSAR modelling has been recently performed for the same data set studied here[12] and, just for the sake of comparison, the leave-10%-out cross-validated Q^2 obtained in that work was 0.767 when using a model in which each compound in each of the 11 cycles is followed by the 11th next compound, and 0.695 when using another model in which each compound in each of the 10 cycles is followed by the10th next compound. This demonstrates that the method presented here is an alternative, suitable QSAR technique.

It is possible to correlate many things through using multivariate regression, but in this study the good correlation did not result from happenstance and to assure that the calibration was not a fortuitous correlation, the Y-block (the activities block) was scrambled and no predictive relationship was found from the modelling (R^2 of 0.415 for three latent variables using model 1, and 0.439 for three latent variables using model 2), as supposed if one considers that a set of compounds with no modelling capability is taken.

Since MIA descriptors are binaries, they do not have a direct physicochemical meaning, though they may be used to drive synthesis, depending on the strategy used to model new compounds with biological activity. Accordingly, a simple procedure is described below to illustrate and exemplify one way of how to achieve potential new drugs. Another way, but that may be applied to little more complex systems having a minimum of similarity, is to build a training set composed by two or more compound classes and then predict the activity of a proposed molecule, which is a miscellaneous of the substructures.

Proposing potential drugs

The strategy utilized here to propose some possible actives is based on an evaluation of which, among the 64 molecules presented, class (thio, sulfinyl or sulfonyl compounds), substitution pattern (2-, 3-, 4-, 2,5- or 3,5-substituted ring) and substituent (among the mono-substituted compounds), possess the highest average activity. A quick screening allows one to conclude that sulfonyl compounds form the class with the largest pIC_{50} average (3.09 ± 0.98). 3,5-Disubstituted compounds are the ones with the highest pIC_{50} values (3.68 ± 0.79), and the most promising substituents are F (2.35 ± 0.30), OMe (2.37 ± 0.60) and Cl (2.53 ± 0.41), the most electronegative ones. Thus, the profile of the proposed drug may be organized as shown in Table 4, which also includes thio and sulfinyl compounds. Note that the 3,5-dichloro derivative is not included in Table 4 because it already pertains to the calibration set, and those compounds with a methoxy group attached to the 5-position of the ring are not contained in the calibration set and, thus, their prediction may be unreliable.

Table 2 Experimental and predicted pIC_{50} values[a] for the 2-amino-6-arylsulfonylbenzonitriles (**35–64**) and their thio (**1–19**) and sulfinyl (**20–34**) congeners, using model 2

Cpd	R	Obs.	Fitted	Pred.	CV	Cpd	R	Obs.	Fitted	Pred.	CV
1	H	1.836	1.652		1.631	**33**	3-Cl, 5-Me	3.495	3.292		3.310
2	2-OMe	2.367	1.892		1.937	**34**	3-OMe, 5-CF_3	2.684	3.143		3.004
3	3-OMe	2.222	2.184		1.948	**35**	H	2.699	2.713		2.639
4	2-Me	1.796	1.401		1.394	**36**	2-OMe	3.222	2.953		2.973
5	3-Me	2.215	1.868		2.121	**37***	3-OMe	3.046		3.224	3.003
6	4-Me	0.939	1.007		1.413	**38**	4-OMe	1.602	2.029		3.079
7	2-Cl	2.387	1.513		1.429	**39**	2-Me	2.638	2.462		2.411
8*	3-Cl	2.131		2.089	2.339	**40***	3-Me	3.398		2.908	3.043
9	2-Br	1.523	1.423		1.477	**41**	4-Me	2.022	1.908		2.119
10*	3-Br	2.292		2.775	3.028	**42**	2-Cl	2.387	2.574		2.607
11	3-F	2.009	1.902		1.962	**43**	3-Cl	3.229	3.171		3.287
12	3-CN	2.762	2.364		1.734	**44**	4-Cl	2.523	2.138		2.270
13*	4-CN	1.359		0.410	0.298	**45**	2-Br	2.301	2.484		2.502
14*	3-CF_3	1.893		1.945	1.457	**46**	3-Br	3.268	3.570		3.855
15	3-NH_2	1.502	1.702		2.192	**47**	4-Br	1.699	1.393		1.996
16	3,5-Me_2	3.367	3.073		3.225	**48**	2-F	2.523	2.597		2.571
17*	3-Cl, 5-Me	2.754		3.043	3.295	**49***	3-F	2.523		2.943	3.058
18	3-OMe, 5-Me	2.699	2.971		2.867	**50***	2-CN	2.268		2.700	2.820
19	3-OMe, 5-CF_3	2.292	2.915		2.950	**51**	3-CN	2.620	3.137		3.122
20*	2-OMe	2.319		2.078	2.131	**52**	4-CN	1.097	1.226		2.240
21	3-OMe	1.796	2.390		2.279	**53**	3-CF_3	2.456	2.739		2.507
22	2-Me	1.032	1.608		1.766	**54**	2,5-Cl_2	3.523	3.157		3.082
23	3-Me	1.534	2.074		2.514	**55***	3,5-Cl_2	4.155		3.803	3.887
24*	4-Me	1.310		1.215	1.453	**56**	3,5-Me_2	5.000	4.156		4.104
25	2-Br	1.407	1.630		1.639	**57***	3-Br, 5-Me	4.699		3.237	3.306
26	3-Br	4.097	2.974		2.528	**58**	3-Cl, 5-Me	4.523	4.147		4.187
27*	4-Br	1.694		0.942	1.206	**59**	3-OMe, 5-Me	4.301	4.055		3.645
28	2-CN	2.409	1.866		1.442	**60**	3-OMe, 5-CF_3	4.046	3.816		3.244
29	3-CN	1.848	2.542		2.515	**61**	3-OH, 5-Me	3.367	3.839		4.004
30	3-CF_3	1.398	2.143		2.057	**62**	3-OCH_2CH_3, 5-Me	4.222	4.267		4.182
31*	3,5-Me_2	3.469		3.274	3.432	**63***	3-$O(CH_2)_2CH_3$, 5-Me	4.222		4.119	4.122
32	2,5-Cl_2	2.007	2.302		2.509	**64**	3-$O(CH_2)_3CH_3$, 5-Me	3.222	3.750		4.001

[a] Fitted values for the training set, predicted values (Pred.) for the test set, and leave-one-out cross-validation values (CV).

Table 3 Statistical parameters[a] of calibration and validation for models 1 and 2

Model	LV	R^2	RMSEC	F	T	Q^2	RMSEP	LV_{cv}	Q^2_{CV}	RMSECV
1	3	0.814	0.479	163.1	44.8	0.823	0.479	3	0.712	0.520
2	3	0.800	0.426	145.4	43.2	0.747	0.548	4	0.624	0.599

[a] LV refers to the number of latent variables utilized, RMSEC, RMSEP and RMSECV are the root mean square errors of calibration, prediction and cross-validation, respectively, and the remaining parameters are defined in the text.

Table 4 Predicted activities (pIC_{50}) for the proposed compounds

	Ring substituent		Model 1[a]			Model 2[b]		
Cpd	3-Position	5-Position	Thio	Sulfinyl	Sulfonyl	Thio	Sulfinyl	Sulfonyl
A	F	F	2.602	2.447	3.452	2.862	3.128	3.889
B	F	Cl	2.597	2.443	3.417	2.942	3.208	3.969
C	F	OMe	2.544	2.420	3.395	2.373	2.639	3.156
D	Cl	F	2.798	2.643	3.618	2.815	3.081	3.842
E	Cl	OMe	2.762	2.586	3.561	2.198	2.464	2.982
F	OMe	F	2.481	2.327	3.302	3.009	3.275	4.036
G	OMe	Cl	2.509	2.354	3.329	3.089	3.355	4.116
H	OMe	OMe	2.435	2.281	3.255	2.496	2.762	3.279

[a] Calibration set of 64 molecules and three latent variables used, [b] calibration set of 64 molecules and four latent variables used.

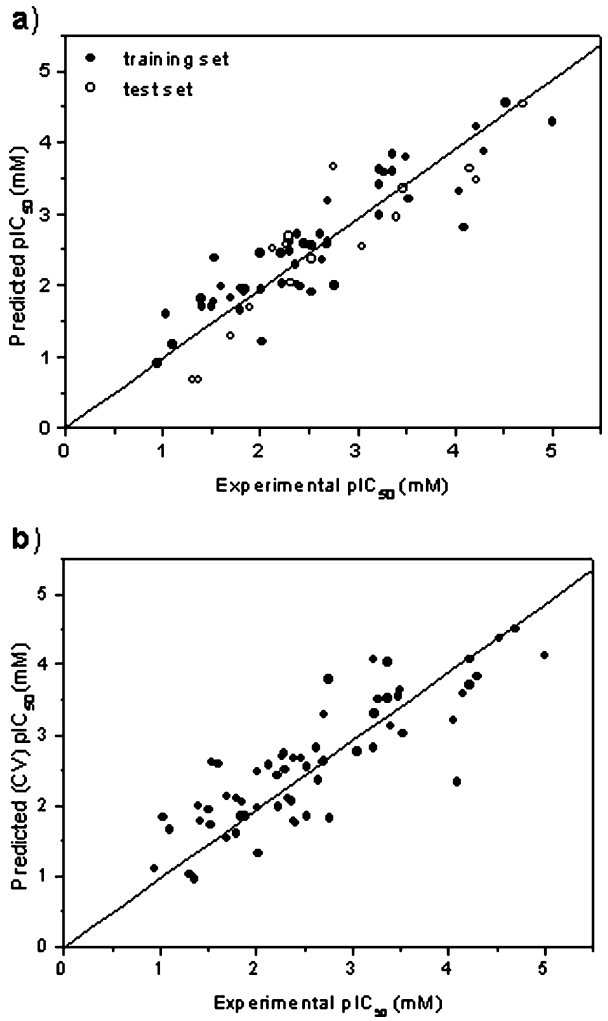

Fig. 3 Plots of experimental *versus* predicted pIC_{50} for the anti-HIV-1 compounds using model 1. (a) Training and test set data and (b) leave-one-out cross-validation data.

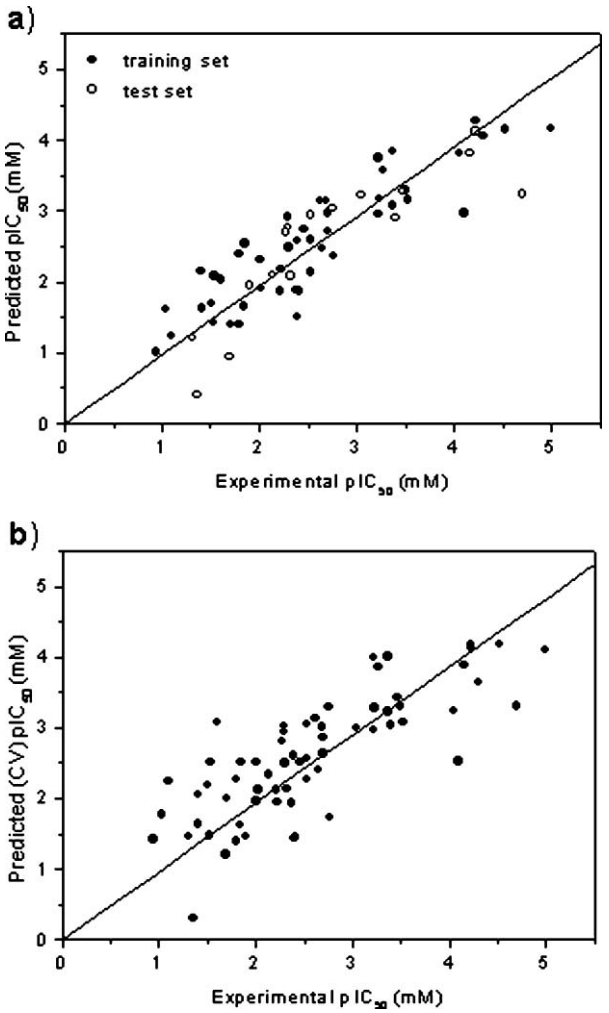

Fig. 4 Plots of experimental *versus* predicted pIC_{50} for the anti-HIV-1 compounds using model 2. (a) Training and test set data and (b) leave-one-out cross-validation data.

According to model 1, the 3-Cl,5-F-sulfonyl compound (**D**-sulfonyl) should present the highest activity among the proposed molecules (predicted pIC_{50} of 3.62), and is also one of the favourites according to model 2. However, this latter model predicts the 3-OMe,5-Cl-sulfonyl compound (**G**-sulfonyl) as the most promising drug. Actually, all of the proposed sulfonyl compounds present reasonably high predicted activities, excepting compounds with a methoxy group bonded to the 5-position of the ring (according to model 2). Thio and sulfinyl compounds showed low predicted activities when compared with the sulfonyl ones, as strategically expected, but they may be useful, since they are less polar than the corresponding sulfonyl compounds and this may affect positively their absorption, as suggested below.

The process of drug discovery requires more than prediction of compounds with high activity. Estimation of molecular transport properties, particularly intestinal absorption and blood-brain barrier penetration is of great interest in the course of a drug development. Traditionally, calculated values of the octanol/water partition coefficient have been used for this purpose and, in the recent years, other parameters have been introduced for absorption prediction.[18] A set of rules imposing limitations on log*P*, molecular weight, and the number of hydrogen bond donors and acceptors, known as "rule of five", was introduced by Lipinski.[19] The rule states that most "drug-like" molecules have log$P \leq 5$, molecular weight $\leq$500, number of hydrogen bond acceptors $\leq$10, and number of hydrogen bond donors $\leq$5. Molecules violating more than one of these rules may have problems with bioavailability. Another very helpful parameter for the prediction of absorption is the topological polar surface area (TPSA), whose correlation with bioavailability data is well illustrated in Fig. 5 for 20 representative drugs extracted from Palm *et al.*[20] Calculations with the Molinspiration program[10] provide suitable parameters to evaluate bioavailability perspectives for the most promising drugs suggested by MIA-QSAR, as shown in Table 5.

Sulfonyl compounds presented larger TPSA values than for the corresponding sulfinyl and thio compounds, as supposed if the increasing in the number of electronegative oxygen atoms causes an increase in the molecular polar surface area. High TPSA may affect drug absorption (see Fig. 5), though sulfonyl compounds have shown to be more active than their congeners and did not violated any of the "rule of five". Compound **56**, which presented the highest measured activity, showed ADME

Table 5 Parameters[a] used for absorption estimation, for selected molecules

Compound	log*P*	*MW*	N_{ON}	$N_{OH,NH}$	N_{rot}	TPSA
A-Thio (3,5-F)	2.738	262.3	2	2	2	49.82
A-Sulfinyl (3,5-F)	1.423	278.3	3	2	2	66.89
A-Sulfonyl (3,5-F)	1.820	294.3	4	2	2	83.96
B-Thio (3-F,5-Cl)	3.347	278.7	2	2	2	49.82
B-Sulfinyl (3-F,5-Cl)	2.032	294.7	3	2	2	66.89
B-Sulfonyl (3-F,5-Cl)	2.430	310.7	4	2	2	83.96
D-Thio (3-F,5-OMe)	3.347	278.7	2	2	2	49.82
D-Sulfinyl (3-F,5-OMe)	2.032	294.7	3	2	2	66.89
D-Sulfonyl (3-F,5-OMe)	2.430	310.7	4	2	2	83.96
F-Thio (3-OMe,5-F)	2.665	274.3	3	2	3	59.05
F-Sulfinyl (3-OMe,5-F)	1.350	290.3	4	2	3	76.12
F-Sulfonyl (3-OMe,5-F)	1.748	306.3	5	2	3	93.19
G-Thio (3-OMe,5-Cl)	3.275	290.8	3	2	3	59.05
G-Sulfinyl (3-OMe,5-Cl)	1.960	306.8	4	2	3	76.12
G-Sulfonyl (3-OMe,5-Cl)	2.357	322.8	5	2	3	93.19
56	2.435	286.3	4	2	2	83.96

[a] log*P* is the octanol/water coefficient partition, *MW* is the molecular weight, N_{ON} is the number of hydrogen bonding acceptors, $N_{OH,NH}$ is the number of hydrogen bonding donors, N_{rot} is the number of rotatable bonds and TPSA is the topological polar surface area.

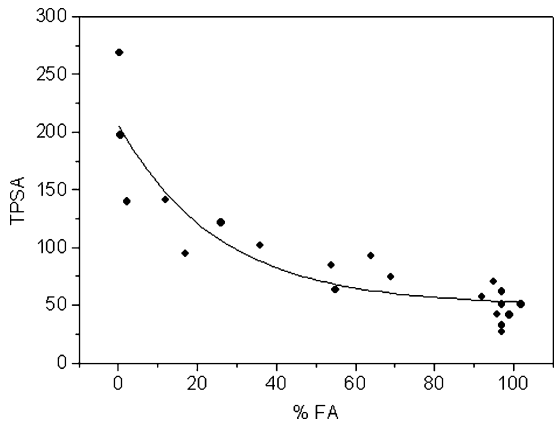

Fig. 5 Percent of drug absorbed after oral administration *vs.* TPSA for 20 molecules from ref. 18.

parameters comparable to the proposed sulfonyl compounds. Which compound should be chosen for a hypothetical synthesis? In this case, it is believed that almost every sulfonyl compound proposed could be a potentially efficient drug, but in general a balance between high activity and good ADME-Tox parameters has to be taken into account.

Conclusions

The goal of this work was to consolidate the MIA-QSAR method as a highly predictive and facile to handle QSAR methodology, whose principles may be used in several fields of research, such as in remote sensing and clinical analysis. This method allows the application of free drawing software and well known multivariate regression algorithms, such as PLS. In addition, it does not require conformational screening and 3D alignment, but only a 2D alignment step, which is simpler and faster than the current three-dimensional procedures. Obviously, this method does not substitute other ligand-based approaches, since some of them may give important information about electrostatic, steric and hydrogen-bonding interactions, but it is an improved approach in many aspects, for instance due to its low cost.

The MIA-QSAR method demonstrated to be a valuable approach for a set of 2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners. Evaluation of models 1 and 2 allows the conclusion that the manner in which molecules were drawn has moderate influence on the modelling. The MIA-QSAR method, together with strategies for the prediction of ADME parameters, might allows one to predict profiles of potential new drugs, *i.e.* compounds with high activity and good absorption.

References

1 R. D. Cramer, III, D. E. Patterson and J. D. Bunce, *J. Am. Chem. Soc.*, 1988, **110**, 5959–5967.
2 G. Klebe, U. Abraham and T. Mietzner, *J. Med. Chem.*, 1994, **37**, 4130–4146.
3 P. J. Goodford, *GRID*, University of Oxford, Oxford, UK, 1995.
4 S. M. Free and J. W. Wilson, *J. Med. Chem.*, 1964, **7**, 395–399.
5 A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav and C. Duraiswami, *J. Am. Chem. Soc.*, 1997, **119**, 10509–10524.
6 S. L. Dixon and K. M. Merz, Jr., *J. Med. Chem.*, 2001, **44**, 3795–3809.
7 M. P. Freitas and J. A. Martins, *Talanta*, 2005, **67**, 182–186.
8 M. Charton, in *Advances in Molecular Structure Research*, ed. M. Charton, JAI Press, Greenwich, 1999, vol. 5, pp. 25–88.
9 M. P. Freitas, S. D. Brown and J. A. Martins, *J. Mol. Struct.*, 2005, **738**, 149–154.
10 Molinspiration Cheminformatics, Bratislava, Slovak Republic (http://www.molinspiration.com).
11 http://www.unaids.org/en/resources/epidemiology.asp.
12 J. T. Leonard and K. Roy, *QSAR Comb. Sci.*, 2004, **23**, 23–35.
13 *ChemWindow version 5.0.1*, SoftShell International Ltd., Grand Junction, CO, USA, 1995–1997.
14 *ACD/ChemSketch version 8.17*, Advanced Chemistry Development Inc., Toronto, Ont., Canada, 1994–2005.
15 H. Wold, in *Multivariate Analysis*, ed. K. R. Krishnaiah, Academic Press, New York, 1966, pp. 391–420.
16 *Matlab Version 6.5.1*, MathWorks Inc., Natick, MA, 1993.
17 J. H. Chan, J. S. Hong, R. N. Hunter, III, G. F. Orr, J. L. Cowan, D. L. Sherman, S. M. Sparks, B. E. Reitter, C. W. Andrews, III, R. J. Hazen, M. St. Clair, L. R. Boone, R. G. Ferris, K. L. Chech, G. B. Roberts, S. A. Short, K. Weaver, R. J. Ott, J. Ren, A. Hopkins, D. I. Stuart and D. K. Stammers, *J. Med. Chem.*, 2001, **44**, 1866–1882.
18 P. Ertl, B. Rohde and P. Selzer, *J. Med. Chem.*, 2000, **43**, 3714–3717.
19 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–25.
20 K. Palm, P. Stenberg, K. Luthman and P. Artursson, *Pharm. Res.*, 1997, **14**, 568–571.