Research Article
AMERICAN JOURNAL OF PHARMACY AND HEALTH RESEARCH
www.ajphr.com
2013, Volume 1, Issue 2
ISSN : 2321–3647(online)

A QSAR Study on cFMS Inhibitors As Potential Anti-Inflammatory Agents: The 2´-Aminoanilide Derivatives

Brij Kishore Sharma[1]*, Pradeep Pilania[1], Prithvi Singh[2], Yashwant[3]
1. *Department of Chemistry, Government College Bundi-323 001, Rajasthan,*
2. *Department of Chemistry, S. K. Government College Sikar-332 001, Rajasthan,*
3. *School of Pharmaceutical Sciences, Lovely Professional University, Jallandhar, Punjab*

ABSTRACT

The cFMS inhibitory activity of 2´-aminoanilide derivatives has been quantitatively analyzed in terms of Dragon descriptors using CP-MLR. The analysis has provided a rational approach for the development of new 2´-aminoanilide derivatives, the cFMS inhibitors, as potential anti-inflammatory agents. The descriptors identified in CP-MLR analysis have highlighted the role of atomic properties in respective lags of 2D-autocorrelations (MATS7m, MATS5v and GATS2p), path/walk ratio 2-Randic shape index (PW2) and Lovasz-Pelikans' leading eigenvalue index (LP1) to explain the biological actions of 2´-aminoanilide derivatives as cFMS inhibitors. Certain structural fragment (C-001) and functionality (nCrHR) in molecular structures have also shown prevalence to optimize the cFMS inhibitory activity of titled compounds. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Keywords: QSAR, 2´-Aminoanilides; cFMS inhibitors, anti-inflammatory agents, combinatorial protocol in multiple linear regression (CP-MLR).

---

*Corresponding Author Email: bksharma_sikar@rediffmail.com
Received 26 April 2013, Accepted 4 May 2013

Please cite this article in press as: Sharma BK. *et al.*, A Qsar Study on CFMS Inhibitors As Potential Anti-Inflammatory Agents: The 2´-Aminoanilide Derivatives. American Journal of Pharmacy & Health Research 2013.

INTRODUCTION

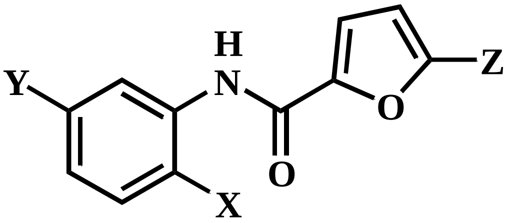
Inflammation plays key role in the progression of human diseases[4,27] The progenitors of inflammation are receiving attention in the potential treatments for diseases. Colony-stimulating factor (CSF-1) or macrophage colony stimulating factor (MCSF) is the primary growth factor for the macrophage lineage[24]. The receptor for CSF-1 or MCSF activation, cFMS, is selectively expressed on macrophages and their progenitor cells. cFMS undergoes dimerization and autophosphorylation upon binding to CSF-1, which ultimately induces the phosphorylation of downstream signaling proteins and leads to differentiation and activation of these cells[15,25]. The over-expression of macrophages at sites of inflammation propagates a chronic inflammatory response through a CSF-1/FMS-driven mechanism[10,29,30]. The activation and proliferation of macrophages is due to the ligation of cFMS by CSF-1. The subsequently release of inflammatory mediators stimulate neighboring fibroblasts to release additional CSF-1 and thereby perpetuate an inflammatory cycle. The cFMS inhibitors are a likely means of interrupting this cycle and result in an anti-inflammatory response[9,5,3,6.] In the activation and proliferation of macrophages, a prominent role played by the CSF-1/cFMS signaling, renders this transduction pathway as an alluring target for therapeutic intervention[13]. Recently, in order to identify selective inhibitors of cFMS as a potential treatment for macrophage-related diseases, a series of 2´-aminoanilides has been reported[14]. In view of the importance of anti-inflammatory agents in the clinical management of several disorders, a quantitative structure-activity relationship is attempted on the cFMS inhibitory activity of these anilides. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavors.

MATERIALS AND METHODS

Chemical structure database and biological activity

This study comprises a chemical structure database of twenty eight 2´-aminoanilide derivatives. These derivatives were assessed in an in vitro enzyme assay for their abilities to inhibit ATP-induced autophosphorylation of cFMS[18]. The structural variations and the cFMS inhibitory activities of titled compounds have been given in Table 1. The reported activity data, IC_{50}(µM), have been converted to the logarithmic scale on molar basis, pIC_{50}(M), and were then used for subsequent QSAR analyses as the response variables. For the purpose of modeling all 28 analogues have been divided into training and test sets. Out of the 28 analogues, one fourth compounds (7) have been placed in the test set for the validation of derived models. The training and test set compounds are also listed in Table 1.

Table1: Structures and observed cFMS inhibitory activity of 2´-aminoanilides.



Cpd.	X	Y	Z	pIC_{50}^{a}
1	Piperidino	H	NO_2	7.28
2	Morpholino	H	NO_2	5.96
3[b]	Pyrrolidino	H	NO_2	5.72
4	Azepino	H	NO_2	6.59
5[b]	2-Methylpiperidino	H	NO_2	6.14
6[b]	3-Methylpiperidino	H	NO_2	6.42
7	3,5-Dimethylpiperidino	H	NO_2	5.28
8	4-Methylpiperidino	H	NO_2	7.11
9	4-Methylpiperazino	H	NO_2	6.41
10[b]	4-Hydroxypiperidino	H	NO_2	6.85
11	4-Hydroxymethylpiperidino	H	NO_2	6.89
12	4-(2-Hydroxyethyl)piperidino	H	NO_2	7.37
13	Piperidino	H	Br	5.36
14	Piperidino	H	Cl	5.72
15[b]	Piperidino	H	CHO	5.96
16	Piperidino	H	CN	7.44
17	Piperidino	OH	CN	7.77
18[b]	Piperidino	OMe	CN	7.05
19	Piperidino	Hydroxymethyl	CN	7.60
20	Piperidino	Aminomethyl	CN	6.70
21[b]	Piperidino	CHO	CN	7.43
22	Piperidino	CN	CN	6.92
23	Piperidino	Piperidin-1-yl	CN	7.77
24	Piperidino	4-Methylpiperazinomethyl	CN	7.10
25	Piperidino	2,3-Dihydroxypropoxymethyl	CN	7.72
26	Piperidino	Methylsulfonamidomethyl	CN	7.43
27	Piperidino	Guanidinylmethyl	CN	7.47
28	4-Methylpiperidino	Hydroxymethyl	CN	7.62

[a]On molar basis (for details see reference Patch *et al.*, 2007); [b]Compounds in test set.

Theoretical molecular descriptors

The structures of the compounds under study have been drawn in 2D ChemDraw. The drawn structures were then converted into 3D modules using the default conversion procedure implemented in the CS Chem3D Ultra. The energy of these 3D-structures was minimized in the MOPAC module using the AM1 procedure for closed shell systems. This will ensure a well defined conformer relationship among the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software[28] for the computation of descriptors for the titled compounds (Table 1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D-, and 2D-descriptor modules. The outlined modules comprised of ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table 2. The combinatorial protocol in multiple linear regression (CP-MLR) procedure has been used in the present work for developing QSAR models[16,17]. Before the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor vs. activity, r < 0.1) were excluded. This has reduced the total dataset of the compounds from 458 to 102 descriptors as relevant ones for the cFMS inhibitory activity. A brief description of the computational procedure is given below.

Table 2: Descriptor classes[a] used along with their definition and scope for modeling the inhibition activity of 2´-aminoanilide derivatives for cFMS inhibitory activity.

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts (MWC)	2D-descriptors representing self-returning walks counts of different lengths
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms
Galvez topological charge indices (GALVEZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
2D-autocorrelations (2D-AUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag)
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups
Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
Empirical (EMP)	1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of

	aromatic bonds and total bonds in an H-depleted molecule
Properties (PROP)	1D-descriptors representing molecular properties of a molecule

[a]Todeschini and Consonni, 2001.

Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies. Its procedural aspects and implementation are discussed in some of our recent publications[19-23]. It involves selected subset regressions. In this procedure a combinatorial strategy with appropriately placed 'filters' has been interfaced with MLR to result in the extraction of diverse structure-activity models, each having unique combination of descriptors from the dataset under study. In this, the contents and number of variables to be evaluated are mixed according to the predefined confines. Here the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3 and upper limit ≤ 0.79). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.71) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated R^2 or Q^2 criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

Model validation

In this study, the data set is divided into training set for model development and test set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F). A number of additional statistical parameters such as the Akaike's information criterion, AIC[1-2] the Kubinyi function, FIT[11], and the Friedman's lack of fit, LOF[7], (Eqs. 1-3) have also been derived to evaluate the best model.

$$\mathrm{AIC} = \frac{\mathrm{RSS} \times (n + p')}{(n - p')^2} \qquad (1)$$

$$\mathrm{FIT} = \frac{r^2 \times (n - k - 1)}{(n + k^2) \times (1 - r^2)} \qquad (2)$$

$$\mathrm{LOF} = \frac{\mathrm{RSS}/n}{\left[1 - \frac{k(d+1)}{n}\right]^2} \qquad (3)$$

where, RSS is the sum of the squared differences between the observed and the estimated activity values, k is the number of variables in the model, p' is the number of adjustable parameters in the model, and d is the smoothing parameter. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value (Fisher ratio), was proved to be a useful parameter for assessing the quality of the models. The main disadvantage of the F-value is its sensitivity to changes in k (the number of variables in the equation, which describe the model), if k is small, and its lower sensitivity if k is large. The FIT criterion has a low sensitivity toward changes in k-values, as long as they are small numbers, and a substantially increasing sensitivity for large k-values. The model that produces the minimum value of AIC and the highest value of FIT is considered potentially the most useful and the best. The LOF takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large numbers of parameters. A minimum LOF value infers that the derived model is statistically sound.

The internal validation of derived model was ascertained through the cross-validated index, Q^2, from leave-one-out and leave-five-out procedures. The LOO method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation has been removed once only. Then one model is developed for each reduced data set, and the response values of the deleted observations are predicted from these models. The squared differences between predicted and actual values are added to give the predictive residual sum of squares, PRESS. In this way, PRESS will contain one contribution from each observation. The cross-validated Q^2_{LOO} value may further be calculated as

$$Q^2_{LOO} = 1 - {}^{PRESS}\!/_{SSY} \qquad (4)$$

where, SSY represents the variance of the observed activities of molecules around the mean value. In leave-five-out procedure, a group of five compounds is randomly kept outside the analysis each time in such a way that all the compounds, for once, become the part of the predictive groups. A value greater than 0.5 of Q^2-index hints toward a reasonable robust model.

The external validation or predictive power of derived model is based on test set compounds. The squared correlation coefficient between the observed and predicted values of compounds from test set, r^2_{Test}, has been calculated as

$$r^2_{Test} = 1 - \frac{\sum (Y_{Pred(Test)} - Y_{(Test)})^2}{\sum (Y_{(Test)} - \bar{Y}_{(Training)})^2} \qquad (5)$$

where, $Y_{Pred(Test)}$ and $Y_{(Test)}$ indicate predicted and observed activity values, respectively of the test-set compounds, and $\bar{Y}_{(Training)}$) indicate mean activity value of the training set. r^2_{Test} is the squared correlation coefficient between the observed and predicted data of the test-set. A value greater than 0.5 of r^2_{Test} suggests that the model obtained from training set has a reliable predictive power.

Y-randomization

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test[26,14] by repeated scrambling of the biological response. The data sets with scrambled response vector have been reassessed by multiple regression analysis (MRA). The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to the unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Applicability domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. A model is valid only within its training domain, and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain is assessed by the leverage values for each compound[8]. A Williams plot (the plot of standardized residuals versus leverage values (*h*) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the applicability domain is established inside a squared area within ±*x* (standard deviations) and a leverage threshold h^*. The threshold h^* is generally fixed at $3(k+1)/n$ (n is the number of training-set compounds, and k is the number of model parameters) whereas $x = 2$ or 3. Prediction must be considered unreliable for compounds with a high leverage value ($h > h^*$). On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

RESULTS AND DISCUSSION

QSAR results

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 7 compounds have been included in the test set for the validation of the models derived from 21 training set compounds. A total number of 102 significant descriptors from 0D-, 1D- and 2D-classes have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in two and three descriptor(s) have been derived successively to achieve the best relationship correlating cFMS inhibitory activity. These models (with 102 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. The selected models in three descriptors are given below.

$$pIC_{50} = 4.843(1.514)\mathrm{MATS5v} + 6.164(1.513)\mathrm{GATS2p} + 0.921(0.192)\mathrm{nCaR} - 4.102$$
$$n = 21,\ r = 0.854,\ s = 0.440,\ F = 15.264,\ Q^2_{LOO} = 0.606,\ Q^2_{L5O} = 0.358,$$
$$r^2_{randY}(sd) = 0.335(0.139),\ AIC = 0.285,\ FIT = 1.526,\ LOF = 0.308,\ r^2_{Test} = 0.725 \quad (6)$$

$$pIC_{50} = 1.912(0.658)\mathrm{IC1} + 4.751(1.534)\mathrm{MATS5v} + 0.971(0.170)\mathrm{C\text{-}028} - 1.470$$
$$n = 21,\ r = 0.847,\ s = 0.450,\ F = 14.410,\ Q^2_{LOO} = 0.610,\ Q^2_{L5O} = 0.657,$$
$$r^2_{randY}(sd) = 0.378(0.131),\ AIC = 0.297,\ FIT = 1.441,\ LOF = 0.321,\ r^2_{Test} = 0.740 \quad (7)$$

$$pIC_{50} = -14.492(4.476)\mathrm{X0Av} + 5.368(1.170)\mathrm{BELv7} + 4.384(1.387)\mathrm{GATS5e} + 5.323$$
$$n = 21,\ r = 0.836,\ s = 0.465,\ F = 13.134,\ Q^2_{LOO} = 0.560,\ Q^2_{L5O} = 0.589,$$
$$r^2_{randY}(sd) = 0.352(0.124),\ AIC = 0.317,\ FIT = 1.313,\ LOF = 0.342,\ r^2_{Test} = 0.521 \quad (8)$$

In above and all follow up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The $r^2_{randY}(sd)$ is the mean random squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models.

The descriptors MATS5v, GATS2p and GATS5e, participated in above models belong to 2D-AUTO class of Dragon descriptors. The 2D autocorrelations are molecular descriptors which describe how a considered property is distributed along a topological molecular structure. The 2D-AUTO descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involve the summations of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighting component in terms of a physicochemical property has been embedded in this descriptor. As a result these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag n (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component – m for mass or v for volume or e for Sanderson electronegativity or p for polarizability – for its computation. It is evinced, from the signs of regression coefficients of the participating descriptors that all of these descriptors, MATS5v, GATS2p and GATS5e, from this class have shown positive influence to the activity. Thus a higher positive value of descriptors MATS5v (Moran autocorrelation of lag 5 weighted by atomic van der Waals' volumes), GATS2p (Geary autocorrelation of lag 2 weighted by atomic polarizabilities) and GATS5e (Geary autocorrelation of lag 5 weighted by atomic Sanderson electronegativities) will be in favor of activity.

The participated descriptors IC1 and X0Av are from the TOPO class of Dragon descriptors. The TOPO class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The descriptor IC1 is information content index of first order neighborhood symmetry and X0Av is average valence connectivity index (chi-0). The descriptor contributed positively to the activity whereas X0Av negatively to the activity. Thus, suggesting that a higher positive value of information content index of first order neighborhood symmetry (IC1) and a lower value of average valence connectivity index (X0Av) would be beneficiary to the activity.

The descriptor BELv7, in above models, is lone representative of BCUT class of Dragon descriptors. The BCUT descriptors are the first 8 highest and the lowest absolute eigenvalues, BEHwk and BELwk, respectively, for the modified Burden adjacency matrix. Here w refers to the atomic property and k to the eigenvalue rank. The ordered sequence of the highest and the lowest eigenvalues reflect upon the relevant aspects of molecular structure, useful for similarity searching. The positive contribution of descriptor BELv7 to the activity advocates that a higher value of this descriptor would enhance the activity.

The other participated descriptors, nCaR and C-028 belong to FUNC and ACF classes of Dragon descriptors, respectively. The functional class descriptors are molecular descriptors based on the counting of the chemical functional groups. Descriptor nCaR represents the number of substituted aromatic carbon in a molecular structure. The positive sign of regression coefficient of this descriptor suggests that more number of substituted aromatic carbon atoms in a molecular structure would be favorable to the activity. Atom centered fragments (ACF descriptors) are simple molecular descriptors defined as the number of specific atom types in a molecule and their calculation is based on the knowledge of the molecular composition and atom connectivities. The ACF class descriptor C-028, representing R-CR-X fragment in a molecular structure, has shown positive influence on the activity suggesting presence of such type of fragment for improved activity.

The three descriptor models could estimate nearly 73 percent variance in observed activity of the compounds. Considering the number of observation in the dataset, models with up to four descriptors were explored. Following are some four-descriptor models for the activity.

$pIC_{50} = -44.276(8.403)MATS7m + 4.072(1.320)MATS5v + 7.214(1.327)GATS2p$

$-0.547(0.155)nCrHR + 43.251$

$n = 21, r = 0.900, s = 0.380, F = 17.084, Q^2_{LOO} = 0.656, Q^2_{L5O} = 0.537,$

$r^2_{randY}(sd) = 0.429(0.145), AIC = 0.235, FIT = 1.847, LOF = 0.287, r^2_{Test} = 0.668$ (9)

$pIC_{50} = -46.809(9.106)MATS7m + 3.055(1.431)MATS5v + 6.242(1.397)GATS2p$

$-0.546(0.183)C\text{-}001 + 46.824$

$n = 21, r = 0.885, s = 0.406, F = 14.434, Q^2_{LOO} = 0.510, Q^2_{L5O} = 0.609,$

r^2_{randY}(sd) = 0.430(0.142), AIC = 0.268, FIT = 1.560, LOF = 0.328, $r^2_{Test} = 0.664$ (10)

pIC_{50} = – 54.500(15.510)PW2 + 30.580(8.643)LP1 + 5.676(1.536)MATS5v

+ 5.760(1.486)GATS2p

n = 21, r = 0.869, s = 0.431, F = 12.356, $Q^2_{LOO} = 0.541$, $Q^2_{L5O} = 0.588$,

r^2_{randY}(sd) = 0.420(0.125), AIC = 0.302, FIT = 1.336, LOF = 0.370, $r^2_{Test} = 0.517$ (11)

These models have accounted for up to 81 percent variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. The values greater than 0.5 of Q^2-index is in accordance to a reasonable robust QSAR model. The pIC_{50} values of training set compounds calculated using Equations (9) to (11) have been included in Table 3. The models (9) to (11) are validated with an external test set of seven compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the predicted activity values are also reported in Table 3. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

Table 3: Observed and modeled cFMS inhibitory activity of 2´-Aminoanilides.

Cpd.	$pIC_{50}{}^a$			
	Obsd.	Cal.		
		Eq. (9)	Eq. (10)	Eq. (11)
1	7.28	6.89	6.78	6.70
2	5.96	5.94	5.94	6.02
3[b]	5.72	6.04	6.07	5.86
4	6.59	7.07	6.89	7.16
5[b]	6.14	6.10	5.97	6.97
6[b]	6.42	6.40	6.34	6.39
7	5.28	5.44	5.54	5.62
8	7.11	7.09	6.90	7.00
9	6.41	6.34	6.31	6.02
10[b]	6.85	6.81	7.18	6.82
11	6.89	7.17	7.51	7.46
12	7.37	7.04	7.41	7.58
13	5.36	5.38	5.47	5.50
14	5.72	6.18	6.16	6.10
15[b]	5.96	6.86	6.79	6.91
16	7.44	6.77	6.70	6.91
17	7.77	7.41	7.35	7.05
18[b]	7.05	6.79	6.86	6.87
19	7.60	7.23	7.24	7.21
20	6.70	7.44	7.45	7.28
21[b]	7.43	6.98	6.99	7.19
22	6.92	6.87	6.86	7.21
23	7.77	7.80	7.80	7.86

24	7.10	7.06	7.06	7.08
25	7.72	7.64	7.65	7.49
26	7.43	7.79	7.66	7.22
27	7.47	7.60	7.53	7.57
28	7.62	7.37	7.31	7.46

[a,b]See foot note under Table 1.

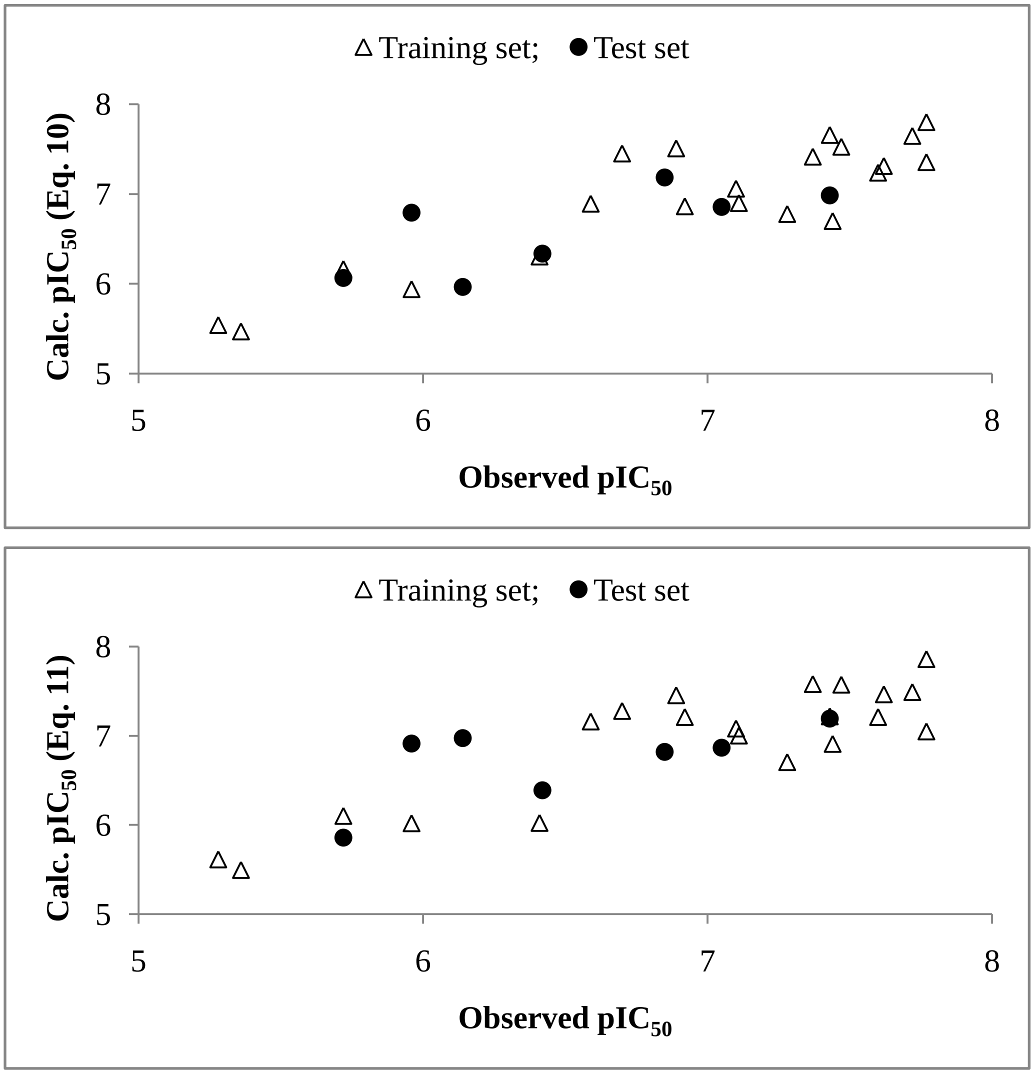


Figure 1. Plot of observed versus calculated activity of training and test set 2´-Aminoanilides.

The newly appeared descriptors in above models are MATS7m (2D-AUTO descriptor), nCrHR (FUNC class descriptor), C-001 (an ACF class descriptor), PW2 and LP1 (TOPO class descriptor). Above Equations reveal that a lower values of Moran autocorrelation of lag 7 weighted by atomic masses (MATS7m), Randic's molecular shape descriptor (PW2, path/walk ratio 2) would be advantageous to enhance the activity. On the other hand, a lower value Lovasz-Pelikan index (LP1) is detrimental to the activity. Counts for certain structural fragment, CH3R and/or CH4 (descriptor C-001), and functionality, sp^3 hybridized ring tertiary carbon atoms (descriptor nCrHR) strongly recommend the absence of such structural features favorable to activity. Thus the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological end point. The QSAR models have further been used to explore some new 2´-aminoanilide derivatives as cFMS inhibitors. The potential structural variations and predicted activity values using model Equations (9-11) for these compounds are given in Table 4.

Table 4: The structures and predicted activity of compounds based on QSAR model Eqs. 9-11.

Cpd.	X	Y	Z	Predicted pIC_{50} based on Model		
				Eq. (9)	Eq. (10)	Eq. (11)
1	Piperidino	Piperidin-1-yl	CHO	7.85	7.81	7.67
2	Piperidino	4-Methyl-piperidin-1-yl	CHO	7.81	7.76	7.99
3	Piperidino	4-Hydroxymethy-piperidin-1-yl	CHO	7.82	8.30	8.29
4	Piperidino	Piperidin-1-yl	NO_2	7.87	7.87	7.88
5	Piperidino	4-Hydroxymethyl-piperidin-1-yl	NO_2	7.80	8.24	815

Applicability domain

On analyzing the model applicability domain (AD) in the Williams plot (Figure 2) of the model based on the whole data set(Table 5), it has appeared that none of the compounds were identified as an obvious outlier for the cFMS inhibitory activity if the limit of normal values for the Y outliers (response outliers) was set as 2.5(standard deviation) units. None of the compounds was found to have leverage (*h*) values greater than the threshold leverages (*h**). For both the training set and test set, the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data. Furthermore, almost all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Table 5: Resultant models for the whole data set (n=28) in descriptors of training set models.

Model	r	s	F	Q^2_{LOO}	Q^2_{L5O}	Eq.
$pIC_{50} = -48.080(7.540)MATS7m + 4.276(1.187)MATS5v + 6.740(1.267)GATS2p - 0.503(0.141)nCrHR + 47.555$	0.883	0.384	20.437	0.699	0.665	9a
$pIC_{50} = -50.209(7.899)MATS7m + 3.258(1.236)MATS5v + 5.701(1.275)GATS2p - 0.483(0.156)C\text{-}001 + 50.790$	0.872	0.402	18.162	0.648	0.665	10a
$pIC_{50} = -54.695(12.783)PW2 + 33.900(8.192)LP1 + 5.895(1.354)MATS5v + 4.921(1.394)GATS2p - 48.493$	0.842	0.442	13.977	0.589	0.604	11a

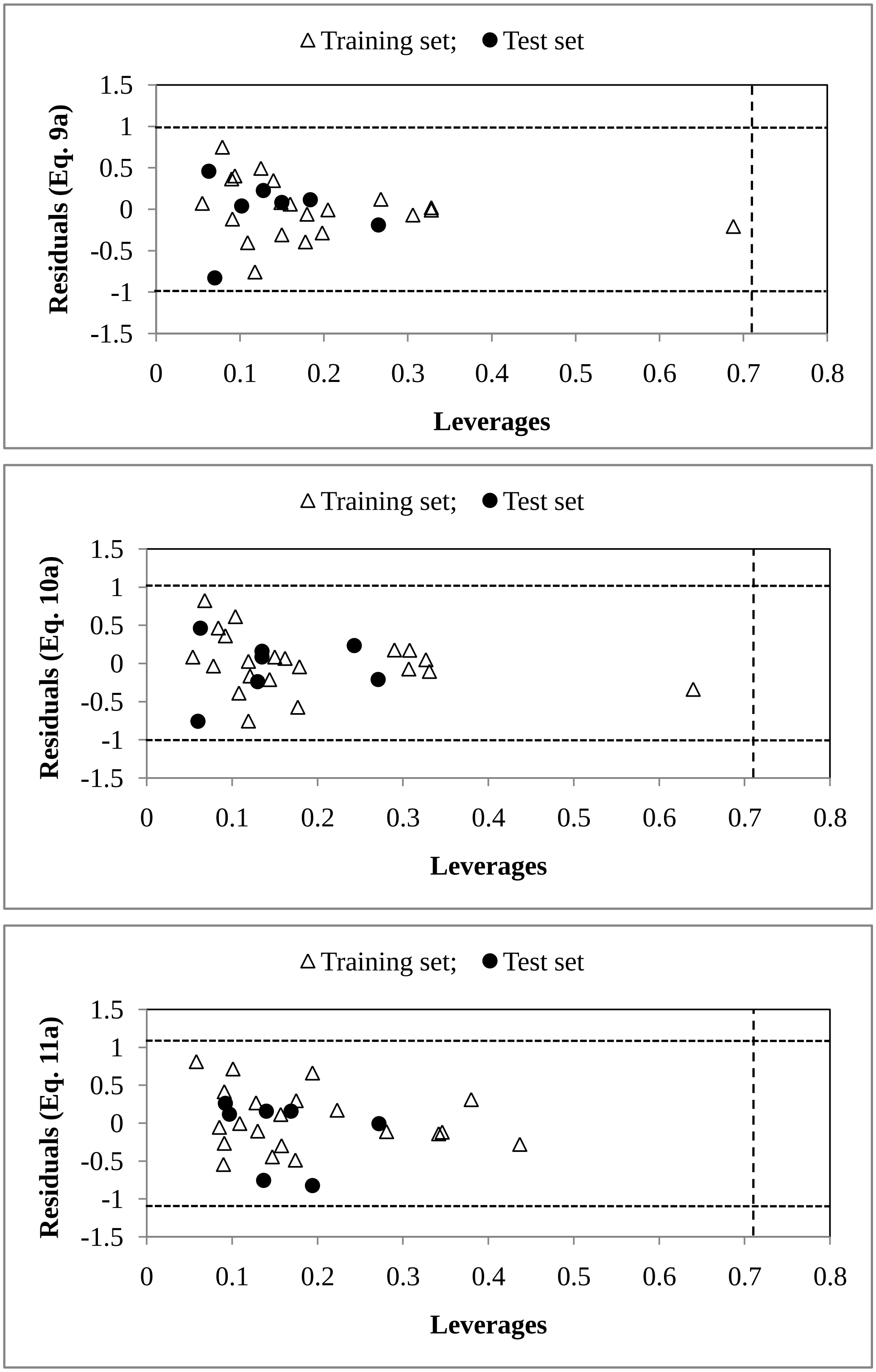


Figure 2. Williams plot for the training set and external prediction set for cFMS inhibitory activity of 2´-Aminoanilide derivatives.

CONCLUSIONS

This study has provided a rational approach for the development of new 2´-aminoanilide derivatives, the cFMS inhibitors, as potential anti-inflammatory agents. The descriptors identified in CP-MLR analysis have highlighted the role of atomic properties in respective lags of 2D-autocorrelations (MATS7m, MATS5v and GATS2p), path/walk ratio 2-Randic shape index (PW2) and Lovasz-Pelikans' leading eigenvalue index (LP1) to explain the biological actions of 2´-aminoanilide derivatives as cFMS inhibitors. Certain structural fragment (C-001) and functionality (nCrHR) in molecular structures have also shown prevalence to optimize the cFMS inhibitory activity of titled compounds. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

ACKNOWLEDGEMENTS

Authors are thankful to their Institutions for providing necessary facilities to complete this study.

REFERENCES

1. Akaike H (1973). Information theory and an extension of the minimum likelihood principle. In: Petrov BN, Csaki F editors. Second international symposium on information theory. Akademiai Kiado, Budapest. pp. 267-281.
2. Akaike H (1974). A new look at the statistical identification model. *IEEE Trans. Automat. Contr.*, AC-19: 716-723.
3. Campbell IK, Rich MJ, Bischof RJ and Hamilton JA (2000). The colony-stimulating factors andcollagen-induced arthritis: exacerbation of disease by M-CSF and G-CSF and requirement for endogenous M-CSF. *J. Leukoc. Biol.*, 68: 144-150.
4. Candore G, Balistreri CR, Grimaldi MP, Vasto S, Listi F, Chiapelli M, Licastro F, Lio D andCaruso C (2006). Age-related inflammatory diseases: role of genetics and gender in thepathophysiology of Alzheimer's disease. *Ann. N.Y. Acad. Sci.*, 1089: 472-486.
5. Conway JG, McDonald B, Parham J, Keith B, Rusnak DW, Shaw E, Jansen M, Lin P, Payne A,Crosby RM, Johnson JH, Frick L, Lin M-HJ, Depee S, Tadepalli S, Votta B, James I, Fuller K, Chambers TJ, Kull FC, Chamberlain SD and Hutchins JT (2005). Inhibition of colony-stimulating-factor-1 signaling in vivo with the orally bioavailable cFMS kinase inhibitor GW2580. *Proc. Natl. Acad. Sci. U.S.A.*, 102: 16078-16083.

6. Cook AD, Braine EL, Campbell IK, Rich MJ and Hamilton JA (2001). Blockade of collagen- induced arthritis post-onset by antibody to granulocyte-macrophage colony-stimulatingfactor (GM-CSF): requirement for GM-CSF in the effector phase of disease. *Arthritis Res.*, 3: 293-298.
7. Friedman J (1990). In: Technical report no. 102. Laboratory for computational statistics. Stanford University, Stanford.
8. Gramatica P (2007). Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.*, 26: 694-701.
9. Hamilton HA (2002). GM-CSF in inflammation and autoimmunity. *Trends Immunol.*, 23: 403-408.
10. Isbel NM, Nikolic-Patterson DJ, Hill PA, Dowling J and Atkins RC (2001). Local macrophageproliferation correlates with increased renal M-CSF expression in human glomerulonephritis. *Nephrol. Dial. Transplant.*, 16:1638-47.
11. Kubinyi H (1994a). Variable selection in QSAR studies. I. An evolutionary algorithm. *Quant. Struct.-Act. Relat.*, 13: 285-294.
12. Kubinyi H (1994b). Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. *Quant. Struct.-Act. Relat.*, 13: 393-401.
13. Lawlor KE, Campbell IK, Metcalf D, O'Donnell K, van Nieuwenhuijze A, Roberts AW and Wicks IP (2004). Critical role for granulocyte colony-stimulating factor in inflammatory arthritis. *Proc. Natl. Acad. Sci. U. S. A.*, 101:11398-11403.
14. Patch RJ, Brandt BM, Asgari D, Baindur N, Chadha NK, Georgiadis T, Cheung WS, PetrouniaIP, Donatelli RR, Chaikin MA and Player MR (2007). Potent 2'-aminoanilide inhibitorsof cFMS as potential anti-inflammatory agents. *Bioorg Med Chem Lett.*, 17: 6070-6074.
15. Pixley FJ and Stanley ER (2004). CSF-1 regulation of the wandering macrophage: complexity in action. *Trends Cell Biol.*, 14: 628-638.
16. Prabhakar YS (2003). A combinatorial approach to the variable selection in multiple linear regression:Analysis of Selwood et al data set – A case study. *QSAR Comb. Sci.*, 22:583-595.
17. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK and Katti SB (2004). CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-Diaryl-1,3- thiazolidin-4-ones. *QSAR Comb. Sci.*, 23: 234-244.
18. Schalk-Hihi C, Ma HC, Struble GT, Bayoumy S, Williams R, Devine E, Petrounia IP, Mezzasalma T, Zeng L, Schubert C, Grasberger B, Springer BA and Deckman, IC

(2006). Protein engineering of the colony-stimulating factor-1 receptor kinase domain for structural studies. *J. Biol. Chem.*, 282: 4085-4093.

19. Sharma S, Prabhakar YS, Singh P and Sharma BK (2008). QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. *Eur. J. Med. Chem.*, 43: 2354-2360.
20. Sharma BK, Pilania P and Singh P (2009). Modeling of cyclooxygnase-2 and 5-lipooxygenase inhibitory activity of apoptosis inducing agents potentially useful in prostate cancer chemotherapy: The derivatives of diarylpyrazole. *J. Enz. Inhibn. Med. Chem.*, 24: 607-615.
21. Sharma S, Sharma BK, Pilania P, Singh P and Prabhakar YS (2009). Modeling of the growth hormone secretagogue receptor antagonistic activity using chemometric tools. *J. Enz. Inhibn. Med. Chem.*, 24: 1024-1033.
22. Sharma BK, Pilania P, Singh P and Prabhakar YS (2010). Combinatorial protocol in multiple linear regression/partial least-squares directed rationale for the caspase-3 inhibition activity of isoquinoline-1,3,4-trione derivatives. *SAR QSAR Environ. Res.*, 21:169-185.
23. Sharma BK, Pilania P, Sarbhai K, Singh P and Prabhakar YS (2010). Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. *Mol. Divers.*, 14: 371-384.
24. Sherr CJ (1988). The fms oncogene. *Biochim. Biophys. Acta.*, 948: 225-243.
25. Sherr CJ, Roussel MF and Rettenmier CW (1988). Colony-stimulating factor-1 receptor (c- fms). *J. Cell Biochem.*, 38: 179-187.
26. So S-S and Karplus M (1997). Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. *J. Med. Chem.*, 40: 4347-4359.
27. Suganami T, Mieda T, Itoh M, Shimoda Y, Kamei Y and Ogawa Y (2007). Attenuation of obesity-induced adipose tissue inflammation in C3H/HeJ mice carrying a toll-like receptor 4 mutation. *Biochem. Biophys. Res. Commun.*, 354: 45-49.
28. Todeschini R and Consonni V (2001) Dragon software (version 1.11-2001). Milano, Italy.
29. Yang N, Isbel NM, Nikolic-Paterson DJ, Li Y, Ye R, Atkins RC and Lan HY (1998). Local macrophage proliferation in human glomerulonephritis. *Kidney Int.*, 54: 143-
30. Yanni G, Whelan A, Feighery C and Bresnihan B (1994). Synovial tissue macrophages and joint erosion in rheumatoid arthritis. *Ann. Rheum. Dis.*, 53: 39-4.