On the Use of the Metric r_m^2 as an Effective Tool for Validation of QSAR Models in Computational Drug Design and Predictive Toxicology

K. Roy* and I. Mitra

Drug Theoretics and Cheminformatics Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Abstract: Validation of quantitative structure-activity relationship (QSAR) models plays a key role for the selection of robust and predictive models that may be employed for further activity prediction of new molecules. Traditionally, QSAR models are validated based on classical metrics for internal (Q^2) and external validation (R^2_{pred}). Recently, it has been shown that for data sets with wide range of the response variable, these traditional metrics tend to achieve high values without truly reflecting absolute differences between the observed and predicted response values, as in both cases the reference for comparison of the predicted residuals is the deviations of the observed values from the training set mean. Roy *et al*. have recently developed a new parameter, modified r^2 (r_m^2), which considers the actual difference between the observed and predicted response data without consideration of training set mean thereby serving as a more stringent measure for assessment of model predictivity compared to the traditional validation parameters (Q^2 and R^2_{pred}). The r_m^2 parameter has three different variants: (i) $r_m^2{}_{(LOO)}$ for internal validation, (ii) $r_m^2{}_{(test)}$ for external validation and (iii) $r_m^2{}_{(overall)}$ for analyzing the overall performance of the developed model considering predictions for both internal and external validation sets. Thus, the r_m^2 metrics strictly judge the ability of a QSAR model to predict the activity/toxicity of untested molecules. The present review provides a survey of the development of different r_m^2 metrics followed by their applications in modeling studies for selection of the best QSAR models in different reports made by several workers.

Keywords: QSAR, r_m^2 metrics, Internal validation, external validation.

INTRODUCTION

The quantitative structure-activity relationship (QSAR) technique plays a crucial role in drug design and ecotoxicological modeling. Remarkable advances in the field of computer sciences over the past decades gave rise to many new possibilities including the ability to simulate and model life's phenomena. Prediction of biological and physical properties of chemical compounds has a long history starting with the linear regression models of Hansch [1] developed during the 1960s. Since then for more than 40 years, QSAR models constitute a major topic in the field of scientific research involving drug design and development and toxicity modeling. The goal of such modeling is to develop statistically acceptable relationship between the molecular structure and biological activity or toxicity or physical property of chemical entities. The molecular properties are calculated based on the molecular structures and are represented as numerical parameters called descriptors [2]. The descriptors encode various molecular features such as the hydrophobic, steric and electronic properties that influence the biological activity or toxicity or property of the molecules under study. These descriptors are then correlated with the corresponding activity/toxicity/property data of the molecules for the development of a QSAR model. With the use of relevant and meaningful descriptors and a statistically significant QSAR model, it is possible to achieve an insight as to what makes a molecule active or inactive. A QSAR model that has been able to capture most of the significant correlations between the structural features and the activity of the molecules may be successfully utilized for activity/property prediction of new molecules belonging to the same class as those used for the development of the QSAR model [3,4]. Thus a predictive QSAR model provides an insight about the structure-activity relationships (SAR) of the molecules used for modeling. Analyses of QSAR models based on relevant descriptors are also useful for improvising necessary structural requirements for activity enhancement or toxicity reduction of the molecules under study as well as for explanation of outliers.

The QSAR technique has been used extensively in a variety of disciplines like drug discovery and lead optimization, risk assessment and toxicity prediction, regulatory decisions and agrochemicals [5-7]. Many thousands of QSAR and QSPR (quantitative structure-property relationship) models have been developed since 1962, covering a wide variety of endpoints and statistical techniques to assess the fitness and validity of significant correlations. Subsequently, guidance for the correct procedures in the development of QSAR/QSPR models has been offered in a number of publications [8-13]. Despite extensive use of QSAR, only recently significant attention has been directed towards validation of the developed QSAR models. Stringent parameters have been employed for extensive validation of the selected QSAR models. The

*Address correspondence to this author at the Drug Theoretics and Cheminformatics Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India; Tel: +91-9831594140; Fax: +91-33-2837-1078; E-mail kunalroy_in@yahoo.com

1875-5607/12 $58.00+.00

validity of a QSAR model is assessed based on four prime tools [14]: (i) cross-validation, (ii) bootstrapping, (iii) randomization of the response data, (iv) external validation by splitting of set of compounds into a training set and a test set followed by confirmation using an independent external validation set or using a designed validation set. In order to have a sound scientific basis for implementation of QSAR models for regulatory use, the REACH (Registration, Evaluation, and Authorization of Chemicals) [15-17] legislation enforced in the European Union inferred that QSARs need to be assessed in terms of their scientific validity. The principles for assessing the validity of QSAR models were proposed at an International workshop held in Setubal (Portugal) in March 2002 and were further modified in 2004 by the Organisation for Economic Co-operation and Development (OECD) Work Programme on QSAR techniques [18].

Recently, there has been a great deal of controversy regarding selection of the most suitable technique for validation of a QSAR model [19]. According to one group of authors [20], internal validation solely serves as a sufficient criterion for selection of the statistically significant QSAR model provided that such internal validation is done properly. According to them, splitting of the dataset performed during external validation may result in loss of information utilized for the development of the model. The model developed based on the reduced training set may lack significant data regarding the molecular properties of the chemical entities. On the contrary, another group of authors state that a QSAR model can only be considered acceptable based on its ability to predict the activity/property of untested molecules. According to Hawkins *et al*. [20], model fit can be best assessed using the crossvalidation technique and it enables to check whether the predictions will carry over to new untested data not used in the model fitting process. Hawkins *et al*. [20] proposed that, in case of a small dataset, holding a portion of the set as test set is insignificant and wasteful. However according to Golbraikh and Tropsha [21], a QSAR model developed using a known set of chemical entities must be validated based on a validation set (test set) of molecules that has not been included for the development of the QSAR model. They proposed several stringent parameters and the QSAR models exceeding the threshold values for each of these parameters can only be considered satisfactory for activity prediction of new molecules.

Viewing the controversy between the internal and external techniques, it is more convenient to apply both of the validation techniques for selection of a statistically significant QSAR model. However, the external predictive parameter (R^2_{pred}) is largely dependent on the selection of the training set compounds. Moreover, the selection of the most significant QSAR model often becomes difficult in cases where comparable models are obtained with different qualities for the internal (Q^2) and external (R^2_{pred}) predictive parameters. An alternative measure r_m^2 was suggested to be a better metric for selection of QSAR models [22-24]. Two variants of r_m^2 have been reported by Roy *et al*. [22, 23]: (i) $r_m^2{}_{(LOO)}$ and $r_m^2{}_{(test)}$. These two parameters are used for judging the internal and external predictivity of the model using the training and test sets respectively. Again, a third parameter, $r_m^2{}_{(overall)}$ has also been developed by Roy *et al*. [23] which may be effectively applied on the whole dataset considering LOO-predicted values for the training set and predicted values of the test set compounds. Thus, this parameter $r_m^2{}_{(overall)}$ analyses the developed QSAR model based on both internal and external validation statistics thereby providing an overall measure of the model predictive ability. Subsequently, selection of the best predictive models from among comparable ones may be performed based on the $r_m^2{}_{(overall)}$ statistic. The r_m^2 metric ($r_m^2{}_{(test)}$ in particular) has been extensively used in different reports for validation purpose. The present review provides an outline of the work involving the development of the r_m^2 metrics followed by its application for statistical validation of QSAR models.

DEVELOPMENT OF THE r_m^2 METRICS

The success of any QSAR model depends on accuracy of the input data collection, selection of suitable descriptors and statistical tools and most importantly the validation of the developed model. Several rules and conditions have been adopted in order to ensure selection of meaningful descriptors from the range of descriptors belonging to different category. According to Topliss, a comparison between the number of variables included in the final model and the number of compounds used for building the model [25] determines the degree of chance correlation, which further increases with an increase in the number of variables. Consequently, the reliability and accuracy of a QSAR model development procedure are established through appropriate validation of the process involved in developing the QSAR model and thus the validation step plays the most important role in the QSAR model building process [26-34]. Hence, the models should be validated both internally and externally in order to check their robustness and predictive potential.

The internal validation procedure involves the leave-one-out (LOO) or leave-many-out (LMO) cross-validation technique followed by the calculation of the cross-validated squared correlation coefficient, LOO-Q^2 or LMO-Q^2 [19, 35]. These techniques involve removal of one or group of compounds from the training set followed by development of the QSAR model based on the reduced dataset. The model thus built with the remaining molecules is used to predict the activity of the deleted compound/compounds. This cycle is repeated till all the molecules of the dataset have been deleted once. The cross-validated squared correlation coefficient (LOO-Q^2 or LSO-Q^2) is calculated according to the following formula (Eq. **1**).

$$Q^2 = 1 - \frac{\sum (Y_{obs(train)} - Y_{pred(train)})^2}{\sum (Y_{obs(train)} - \overline{Y}_{training})^2} \quad (1)$$

In the above equation, $Y_{obs(train)}$ is the observed activity (training set), $Y_{pred(train)}$ is the predicted activity of the training set molecules based on the LOO/LMO technique while $\overline{Y}_{training}$ is the mean activity data of the training set compounds. A model is considered to be satisfactory if the value of Q^2 exceeds the stipulated value of 0.5. Moreover, although a high value of LOO-Q^2 appears to be a necessary criterion but it is not sufficient enough for ensuring model

predictivity [36]. A problem with LOO cross-validation is that a small change in the data can cause a huge variation in the type of QSAR model selected [37]. Thus, a QSAR or QSPR model is chiefly valued in terms of its predictivity, indicating its ability to predict the response parameters for compounds not used in developing the correlation, i.e. molecules not included in the training set. Such a procedure for checking model predictivity based on molecules not included in the training set is referred to as external validation.

In many cases, truly external data points become unavailable for prediction purpose; thus, the original dataset compounds are split into training and test sets where the training set is used for building the QSAR model and the corresponding test set is used for its subsequent validation. The test set may vary in size depending upon the total number of compounds in the original dataset [38]. Thus proper splitting of the dataset into the corresponding training and test sets plays the prime role for the successful development of a QSAR model. The training set should be selected in such a way that they represent the entire class of compounds from which they originate both in terms of the chemical features and activity profile. The QSAR model thus developed is used for activity prediction of the test set molecules followed by the estimation of the external predictive parameter (R^2_{pred}) (Eq. **2**) [21] which reflects the degree of correlation between the observed and predicted activity data for the test set molecules, thereby ensuring the model predictive ability.

$$R^2_{pred} = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{\sum (Y_{obs(test)} - \overline{Y}_{training})^2} \quad (2)$$

In Eq. (**2**), $Y_{obs(test)}$ and $Y_{pred(test)}$ are the observed and predicted activity data respectively for the test set compounds. Models with values of R^2_{pred} above the threshold value of 0.5 are considered to be well predictive.

From the above equations, it can be noted that the values of Q^2 and R^2_{pred} are dependent on the mean activity value of the training set compounds and its distance from each of the activity values of the corresponding training and test set compounds respectively. As the denominator term in both the equations increases [$\sum (Y_{observed} - \overline{Y}_{training})^2$], the values of the internal and external predictive parameters increase, apparently suggesting improved predictive ability of the developed QSAR model. Thus, a dataset comprising of molecules exhibiting a wide activity range may show significantly acceptable values for the two parameters, although large differences may exist between the predicted and corresponding observed activity values for the training and test set molecules. Hence, a large value of Q^2 or R^2_{pred} does not necessarily indicate that the predicted and the observed activity data are in close proximity to each other, although there may exist a good overall correlation between the values. Thus, to better indicate both the internal and external predictive capacities of a QSAR model, a modified r^2 term (r_m^2) [22-24] has been developed by Roy *et al*.

$$r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (3)$$

In Eq. (**3**), r^2 and r_0^2 are the squared correlation coefficient values between the observed and predicted activity data with and without intercept respectively. Models with values of the r_m^2 metrics above the threshold value of 0.5 are considered to be good predictive ones. For good prediction, the predicted activity values (for both training and test sets) should be close to the corresponding observed activity data. This results in a decrease in the difference between the values of r^2 and r_0^2. Thus, for the ideal case (best prediction), r^2 becomes equal to r_0^2 and accordingly, r_m^2 equals to r^2. On the contrary, for the worst model the value of r_m^2 tends to zero (or even negative some time). In case of internal validation, the r_m^2 term is referred to as $r_m^2{}_{(LOO)}$ and is calculated based on the observed and LOO predicted activities of the training set compounds. Similarly, in case of external validation, the parameter is referred to as $r_m^2{}_{(test)}$ and is determined using the observed and predicted activity data of the test set molecules as calculated from the developed QSAR model. It can be inferred from the above equation that the parameters $r_m^2{}_{(LOO)}$ and $r_m^2{}_{(test)}$ solely depend on the observed and the predicted activity data of the training and test set compounds respectively and any significant difference between the observed and predicted activity values is well reflected in values of the r_m^2 metrics. Another variation of the r_m^2 metric includes the calculation of the 'true $r_m^2{}_{(LOO)}$' parameter based on the model developed from the undivided data set after the application of variable selection strategy at each cycle of validation. It has been shown by Mitra *et al*. [24] that for a small dataset, the 'true $r_m^2{}_{(LOO)}$' parameter reflects characteristics of external validation for the developed QSAR model. In such cases, the parameter may be utilized for determining the predictive ability and the accuracy of prediction of the developed QSAR model.

Moreover, the r_m^2 statistic is not only applied for prediction for the training and test sets individually, it can also be extended to the entire dataset combining both the training and test sets. According to Roy and coworkers [39], this metric can be calculated based on the entire dataset considering the LOO predicted activity for the training set and predicted activity for the test set compounds. Thus, this parameter refers to the overall predictive potential of the entire dataset and is represented as $r_m^2{}_{(overall)}$. The use of $r_m^2{}_{(overall)}$ statistic as a metric for selection of a statistically significant QSAR model has two prime advantages. Firstly, the $r_m^2{}_{(overall)}$ metric includes the entire dataset including the training and test sets. Thus, unlike internal and external validation metrics which are derived from reduced datasets, the $r_m^2{}_{(overall)}$ performs prediction based on comparably larger number of compounds thereby imparting greater reliability to the assessment of prediction capacity of the QSAR model. Additionally, often comparable models are obtained in terms of the internal and external validation parameters making it difficult to select the best one. Since the $r_m^2{}_{(overall)}$ parameter involves the whole dataset under study, it refers to the overall contributions of predictions for both the internal and external validation sets. It thus enables selection of the most significant QSAR model based on the predictive ability of

the model for the whole dataset. Finally, it may be inferred that the parameters $r_m^2{}_{(LOO)}$ and $r_m^2{}_{(test)}$, being independent of the range of activity data of the compounds under study, serve as stricter metrics than Q^2 and R^2_{pred} for internal and external validations respectively. Moreover, the parameter $r_m^2{}_{(overall)}$ is unique in terms of its ability to estimate the overall predictive ability of the model for the class of compounds used for the study irrespective of their classification into training and test sets.

APPLICATION OF THE r_m^2 METRICS FOR THE VALIDATION OF STATISTICALLY SIGNIFICANT QSAR MODELS

The r_m^2 metrics have been used widely by several authors in the recent QSAR literature to check predictive ability of the developed QSAR models. The $r_m^2{}_{(test)}$ metric analyses the models chiefly based on their ability to predict the activity of the test set compounds, without any consideration of training set mean and thus it enables better screening of the most significant QSAR model. This metric was introduced by Roy *et al*. [22] in 2008. The authors explored the optimum variable selection strategy for Partial Least Squares (PLS) regression technique using a model dataset of cytoprotection data and finally ended up with the development of a new parameter, $r_m^2{}_{(test)}$, for checking external predictability of the QSAR models. The work has been performed by dividing the whole dataset into ten different combinations of training and test sets using the *k*-means clustering technique followed by the development of PLS models. The models were developed with number of PLS components optimized by the LOO-Q^2 and were subsequently validated (externally) using the corresponding test set compounds. From the study, it was concluded that model performance chiefly depends on the type of training set selected for fitting the model. Moreover, according to this work, the parameter R^2_{pred} should not be considered as the ultimate criterion for indicating the external predictability of a model since a clear separation between the curves of r^2 and r_0^2 in some trials suggested that such models might not be truly predictive in spite of bearing acceptable values of R^2_{pred}. Thus, to truly explore the predictive potential of a QSAR model, the difference in the values of r^2 and r_0^2 should be small and the extent of this difference is reflected in the value of the $r_m^2{}_{(test)}$ parameter, the threshold of 0.5 for which is considered acceptable.

In another paper, Roy *et al*. [39] introduced the concepts of $r_m^2{}_{(LOO)}$ and $r_m^2{}_{(overall)}$ parameter. In a comparative QSAR study of CYP1A2 inhibitor flavonoids using 2D and 3D descriptors, the authors employed the r_m^2 metrics for the selection of the final statistically significant QSAR models. This work reveals that the concept r_m^2 may be well extended beyond only test set prediction and can be utilized for training set prediction if one considers the correlation between observed and LOO-predicted values. Moreover, the concept can also be used for the whole set considering LOO-predicted values for the training set and predicted values of the test set compounds. Thus, the G/PLS model developed using the 2D descriptors, among all the developed models, exhibited maximum values for these three parameters [$r_m^2{}_{(test)}$ = 0.685, $r_m^2{}_{(LOO)}$ = 0.736, $r_m^2{}_{(overall)}$ = 0.695] and thus, it was selected as the best one. On the contrary, although models developed with the 3D descriptors exhibited statistically acceptable values for the Q^2 and R^2_{pred} metrics, these were considered as unreliable models as none of them were able to attain the stipulated value for the $r_m^2{}_{(overall)}$ parameter. The r_m^2 metrics thus indicate the predictive ability of the model both in terms of internal and external validation measures and the results obtained in the work indicate a close correspondence between the observed and corresponding predicted activity data for both the training and test sets. However, it is to be noted that in cases where more number of compounds are included in the modeling set while splitting of a dataset, the value of $r_m^2{}_{(overall)}$ leans towards the efficiency of the model for activity prediction of the training set molecules to a greater extent compared to that for the test set molecules.

In yet another paper, Roy *et al*. [23] elaborated the concepts of the r_m^2 metrics based on three different datasets each of which was divided into 50 different combinations of training and test sets, and each model thus obtained was validated based on internal and external validation measures. In case of the results obtained for the first dataset, it was observed that though some of models exceeded the stipulated values of the traditional validation parameters, but none of them satisfied the value of $r_m^2{}_{(overall)}$. This may be explained by the fact that high values of Q^2 and R^2_{pred} are obtained as long as a good correlation exists between the observed and predicted activity data of the training and test sets respectively but this does not necessarily mean that the predicted values are very close to the corresponding observed ones. The $r_m^2{}_{(overall)}$ statistic, being entirely dependent on the observed and predicted activity data, rightly reflects a difference between the observed and predicted activity values. Again in case of the 2nd dataset, due to the wide distribution of the ovicidal activity among the congeners, acceptable values of the two parameters, Q^2 and R^2_{pred} were obtained in spite of considerable differences in numerical values of the observed and predicted activities for some compounds. Due to large differences in the values of the observed and predicted activity data, none of the models could attain the threshold value for the $r_m^2{}_{(overall)}$ parameter and thus these were considered to be non-reliable. The results obtained in this work reveal that the values of Q^2 and R^2_{pred} may differ considerably for a given model, i.e., a model with a high value of Q^2 may show a low value for R^2_{pred} and vice versa. The $r_m^2{}_{(overall)}$ metric takes into consideration both the internal and external validation measures and penalizes a model for a difference in the values of the two parameters. Again in case of the 3rd dataset, most of the models are acceptable in terms of all the traditional validation parameters as well as those developed by Roy *et al*. [22]. Thus in such a case, the selection of the best model based on either internal or external predictive parameters becomes difficult. Consequently, the $r_m^2{}_{(overall)}$ parameter which takes into consideration the prediction for both the training and test sets have been aptly used as the measure for QSAR model selection.

Roy and Popelier [40] made interesting observations in a work involving exploration of predictive QSAR models for hepatocyte toxicity of phenols using quantum topological molecular similarity (QTMS) descriptors. The QTMS descriptors were calculated at different levels of theory which include AM1, HF/3-21G(d), HF/6-31G(d), B3LYP/6-

31+G(d,p), B3LYP/6-311+G(2d,p) and MP2/6-311+G(2d,p) followed by the development of models based on these descriptors at each level. It was observed that the best model at each level was distinguished from others by a close proximity between the values of the two external predictive parameters, R^2_{pred} and $r_m^2{}_{(test)}$. Thus, it may be inferred that for models having such a close correspondence between the two values of R^2_{pred} and $r_m^2{}_{(test)}$, the activity data predicted based on the corresponding models closely match with the observed activity values. In another work Roy and Popelier [41] also developed predictive QSAR models using QTMS descriptors for determining the toxicity of nitroaromatics to *Saccharomyces cerevisiae*. The results were analysed based on different internal and external validation techniques. However selection of the best model was difficult based on the maximum values of Q^2 (0.910) and R^2_{pred} (0.971) because the model with maximum Q^2 value yielded lower value for the R^2_{pred} parameter and vice versa. Thus, the value of $r_m^2{}_{(test)}$ was calculated for all the models so as to enable the selection of the best developed QSAR model. Based on the results of validation and randomization, three best models (model nos. 1, 7 and 9) were selected to be statistically significant ones. It was revealed that the toxicity data predicted for the test set molecules using model nos. 1 and 9 were close to the observed toxicity values. This observation is well reflected in the closely coinciding values of the two external predictive parameters, R^2_{pred} and $r_m^2{}_{(test)}$ for model nos. 1 and 9. However, in case of model no. 7, the toxicity predicted for most of the test set compounds was close to the corresponding observed data with the exception of compound no. **11.** This difference in the observed and predicted toxicity value for compound no. **11** remains unnoticed during the calculation of the R^2_{pred} parameter (which exhibits a maximum value of 0.932 for model no. 7 among all the acceptable models developed) but is well reflected in the reduced value of the $r_m^2{}_{(test)}$ parameter (0.849). Thus, it may be inferred that due to its ability to analyse the deviation between the observed and predicted activity data of the test set molecules more precisely, the $r_m^2{}_{(test)}$ parameter is a stricter parameter for measuring external predictive ability of a QSAR model compared to R^2_{pred}. Again, the $r_m^2{}_{(test)}$ metric has been aptly used in developing predictive QSPR modeling of the acidic dissociation constant (pKa) of phenols in different solvents by Roy and Popelier [42] based on QTMS descriptors calculated at different levels of theory. The work once again suggested that the R^2_{pred} parameter should not be solely used as the ultimate criterion for reporting the external predictive ability of a QSAR model. Amongst all the developed models, the model with maximum R^2_{pred} (0.968) value yielded a reduced value for the $r_m^2{}_{(test)}$ (0.916) parameter while that with maximum $r_m^2{}_{(test)}$ (0.986) value also exhibited an appreciably acceptable value for the R^2_{pred} (0.948) parameter. Thus the authors used both the parameters for judging the external predictivity of the QSAR models and the best model for each solvent was selected based on the value of $r_m^2{}_{(test)}$.

Several papers reported by Roy *et al.* utilize the r_m^2 metrics for the selection of the best QSAR models. Roy *et al*. [43] explored 2D and 3D QSARs of 2,4-diphenyl-1,3-oxazolines for ovicidal activity against *Tetranychus urticae* and employed different validation measures for the selection of the best QSAR model. An analysis of the results obtained from the validation of the different models revealed that although a maximum value of R^2_{pred} (0.755) was obtained for the genetic function approximation with spline option (GFA-spline) model (model 6) developed using the shape, spatial, electronic and physicochemical descriptors, the authors reported model 4 as the most significant one (GFA-spline model developed using the topological, structural and physicochemical descriptors). Such a selection of the best model was performed based on the $r_m^2{}_{(overall)}$ parameter. Despite bearing the maximum value for the R^2_{pred} (0.755) parameter, model 6 exhibits a lower value for the $r_m^2{}_{(overall)}$ (0.526) parameter. On the contrary, model 4 yielding the maximum value for the $r_m^2{}_{(overall)}$ (0.535) parameter was selected as the most significant QSAR model. As the $r_m^2{}_{(overall)}$ metric takes into consideration the predicted and observed activity data of both the training and test sets and does not depend on the range of activity data like R^2_{pred}, it serves as a better measure for the selection of the best QSAR model from among comparable ones.

In a study involving the development of comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) models of cytotoxicity data of anti-HIV-1-phenylamino-1H-imidazole derivatives by Basu *et al*. [44], the authors developed the models and validated them using 10% and 25% of molecules as test set, based on the Database and Field-Fit alignment techniques. The authors reported the best QSAR (CoMFA analysis using database alignment and employing 25% molecules in the test set) based on the values of both R^2_{pred} and r_m^2 parameters (R^2_{pred} of 0.91 and r_m^2 of 0.652). Although the CoMSIA model obtained using the same test set also yielded a statistically significant value of the external predictive parameter (R^2_{pred}=0.789), it could not reach the threshold for the modified r^2 (r_m^2) (0.410) parameter because some of the compounds show a marked difference in the values of their observed and predicted activity data. Such a difference is well reflected in the calculation of r_m^2 but is not considered in case of R^2_{pred} calculation since the observed and predicted activity data maintains a good overall correlation between them despite having a significant difference in the respective values. Thus, according to these authors, the r_m^2 metric serves as a helpful technique in identifying the best model from amongst many comparable models.

In another work, Roy *et al*. [45] dealt with comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using various chemometric techniques. In this work, Roy *et al*. developed different QSAR models which varied in their degree of acceptability based on the values of Q^2 and R^2_{pred}. The model developed using the stepwise-MLR technique showed maximum predictivity in terms of the external predictive parameter (R^2_{pred} = 0.701), while the model developed using the GFA technique yielded maximum value for the internal predictive parameter (Q^2 = 0.836). Thus, the selection of the best QSAR model developed for activity prediction becomes difficult. This problem was handled by using calculations based on the r_m^2 metrics. The results obtained for the r_m^2

metrics revealed that none of the above mentioned two models could achieve high values for these parameters due to a significant difference between the observed and corresponding predicted activity data for the training and test sets. This is not taken into consideration in the calculations for Q^2 and R^2_{pred} as long as the observed and predicted data maintain an overall good correlation among themselves. On the contrary, the model developed using the genetic partial least squares (G/PLS) technique, although bearing lower values for the Q^2 (0.827) and R^2_{pred} (0.600) parameters, exhibited maximum values for all the r_m^2 metrics [$r_m^2{}_{(LOO)}$ = 0.771, $r_m^2{}_{(test)}$ = 0.581, $r_m^2{}_{(overall)}$ = 0.735] due to a close proximity between the observed and predicted activity data for the training and test set compounds and thus the model was reported as the best model for the work in terms of its overall predictive ability. Another work by Roy *et al*. [46] explains the QSAR and quantitative structure activity-activity relationship (QAAR) studies for inhibitors of cytochrome P450 2A6 and 2A5 enzymes using GFA and G/PLS techniques. The results obtained in this study once again imply the significance of the $r_m^2{}_{(test)}$ parameter. An analysis of model 5 developed in the reported work reveals that the model was selected as the best one among all based on its internal and external predictive ability. A significant difference between the values of the R^2_{pred} and $r_m^2{}_{(test)}$ parameters reflects that the parameter R^2_{pred} overlooks a significant difference between the values of the observed and corresponding predicted activity data of the test set molecules as long a good overall correlation is maintained among them, but the $r_m^2{}_{(test)}$ parameter takes into account such difference.

Hu *et al*. [47] employed a variety of chemometric tools performing QSAR studies for a series of HIV-1 reverse transcriptase inhibitors (2-amino-6-arylsulfonylbenzonitriles and their thio and sulfinyl congeners) using topological, geometrical, quantum mechanical energy-related and charge distribution-related descriptors. Among the compounds used for the study, only 64 compounds exhibited IC_{50} (50% inhibitory concentration) values for anti-HIV-1 activity while the remaining 51 compounds were provided with IC_{50} values for HIV-1 RT binding affinity. Both the activity data were modeled separately and the QSAR model obtained with the anti-HIV-1 activity using the projection pursuit regression (PPR) technique yielded maximum values of the squared correlation coefficients (R^2) for both the training (0.890) and test (0.882) set compounds. Subsequent validation of the model based on the r_m^2 statistic also yield maximum value for the $r_m^2{}_{(test)}$ (0.660) parameter among all the models developed with the same series of activity data. Among the predicted activity data computed for the test set molecules using this model, compound nos. **53**, **58** and **65** exhibit a comparatively higher range of residual activity which accounts for a reduction in the value of the $r_m^2{}_{(test)}$ parameter in comparison to that of the R^2_{pred} parameter (yielding a large value of R^2_{pred} due to the wide range of activity data of the compounds under study).

In several reports, Mitra *et al*. employed the parameter $r_m^2{}_{(LOO)}$ as a measure for judging the predictive ability of the developed QSAR models. While analyzing QSAR models for the antioxidant activity of hydroxybenzalacetones, Mitra *et al*. [48] was unable to perform external validation due to lack of enough data for splitting the dataset into training and test sets. Thus, they employed various internal validation techniques for assessing the model predictive ability which include leave-one out and leave-many-out cross validation techniques together with the calculation of the $r_m^2{}_{(LOO)}$ statistic. A high value of the $r_m^2{}_{(LOO)}$ metric increases the reliability of the developed model since it determines the extent of deviation of the predicted activity data from the corresponding observed activity. Among the three end points modeled in the work using different chemometric tools (primarily GFA and G/PLS), the best models were selected based on the $r_m^2{}_{(LOO)}$ parameter. Another work by Mitra *et al*. [49] deals with QSAR studies of antilipid peroxidative activity of substituted benzodioxoles using chemometric tools. In this work also, several models were developed using different chemometric tools. It was observed that the model developed using the G/PLS technique based on the charge and physicochemical descriptors yielded much lower value for $r_m^2{}_{(LOO)}$ (0.704) parameter although it shows a high value for LOO-Q^2 (0.825). However a high value of $r_m^2{}_{(LOO)}$ [$r_m^2{}_{(LOO)}$ = 0.823, Q^2 = 0.784] was obtained for the GFA model developed with the MSA and spatial descriptors. It may be explained by the fact that among the two models, the residual values of compound nos. **11**, **12** and **15** in case of the former model are much higher compared to those of the latter one. Thus, such a difference in the observed and predicted activity data is well reflected in the reduced value of $r_m^2{}_{(LOO)}$ (0.704) for the former model. The latter model having the maximum $r_m^2{}_{(LOO)}$ value was selected as the best one for antioxidant activity prediction of benzodioxoles.

A recent work by Srivastava *et al*. [50] reports the quantitative structure–activity relationship of artemisinin derivatives leading to the development of predictive *in vivo* antimalarial activity models. In this work, the authors quantified the predictive ability of the best QSAR model based on an analysis of the r_m^2 metric. They utilized several types of descriptors including topological, spatial, thermodynamic, information content, lead likeness, and E-state indices and derived a final quantitative relationship between antimalarial activity and structural properties based on stringent internal and external validation parameters. However, the best reported QSAR model showed a marked difference in the values of R^2_{pred} (0.876) and $r_m^2{}_{(test)}$ metrics (0.788). Such a variation in the values of the two parameters may be attributed to the increased residuals for the test set compounds, especially in case of compound nos. **34**, **58**, **137**, **144** and **161** which is well reflected in the calculation of $r_m^2{}_{(test)}$. Thus, the $r_m^2{}_{(test)}$ metric plays a prime role in selecting the best QSAR model with efficient predictive ability

In an estimation of predictive ability of the models for bioconcentration factor developed using SMILES-based optimal descriptors Toporov *et al*. [51] applied the r_m^2 statistics for selection of the final best model. They obtained an $r_m^2{}_{(test)}$ value of 0.657 while an R^2_{pred} value of 0.797 for a developed QSAR model and reported it to be the most satisfactory one. A lower value of $r_m^2{}_{(test)}$ compared to the R^2_{pred} parameter occurs due to high predicted residual values for some compounds which remains unnoticed in case of calculation of the R^2_{pred} parameter.

Nargotra *et al*. reported two different papers related to the development of QSAR models of piperine analogs for bacterial NorA efflux pump inhibitors [52] and the development of QSAR of aryl alkenyl amides/imines for bacterial efflux pump inhibitors [53]. In both the cases, the selection of the final most significant QSAR model was made based on the values of both R^2_{pred} and r_m^2. The models yielded acceptable values for both parameters indicating the statistical significance of both the developed QSAR models. In the two reports mentioned here, the activity of the test set compounds predicted using the developed QSAR models lie in close vicinity to the corresponding observed activity data. Thus in both the reports the two parameters (R^2_{pred} and r_m^2) are close to each other with very little numerical difference between themselves.

Again, while performing docking and 3D QSAR studies of protoporphyrinogen oxidase inhibitor 3H-pyrazolo [3,4-d][1,2,3]triazin-4-one derivatives, Roy *et al*. [54] utilized the $r_{m\ (overall)}^2$ metric as a crucial criterion for the selection of the best QSAR model. A wide distribution of the activity range of the compounds resulted in acceptable Q^2 and R^2_{pred} values for most of the models developed, with model M2 (modified to model M2a due to intercorrelation among descriptors) exhibiting a maximum value for the R^2_{pred} (0.788) parameter. A number of compounds (compound nos. **3**, **29**, **31**) predicted using model M2a exhibit a significant variation between the observed and predicted activity values, but this deviation is not reflected in the value of R^2_{pred}. Thus, the R^2_{pred} parameter is unable to explore the error in the results due to significant deviations between the observed and predicted activity data. Hence, the r_m^2 statistic was utilized as a prime measure for the selection of the most significant QSAR model. Based on the value of $r_{m\ (overall)}^2$ (0.771) parameter, model M3 were selected as the best one despite showing a reduced value for the external predictive R^2 (0.726). Moreover, the residuals calculated for the training and test set compounds using model M3 were much lower compared to those calculated using the remaining models, thereby reflecting the reliability of the selected QSAR model as the best one.

In a study analyzing the binding conformations and QSAR of combretastatin A4 (CA-4) analogs as tubulin inhibitors, Liao *et al*. [55] employed all the three r_m^2 metrics for assessing the predictive ability of the developed QSAR model. Although the values of Q^2 (0.666) and $r_{m\ (LOO)}^2$ (0.654) were close to each other, a marked difference was observed between the values of R^2 for the test set (0.806) and $r_{m\ (test)}^2$ (0.731). This may be explained by the fact that for the training set compounds, the observed and the predicted activity data lie in close proximity to each other. However, in case of the test set, some of the compounds exhibit comparatively larger residual values which ultimately contributes to a lower value of $r_{m\ (test)}^2$. Moreover, the $r_{m\ (overall)}^2$ (0.713) parameter explains the overall predictive ability of the QSAR model (in terms of both training and test sets) and makes a balance between the values of the internal and external predictive parameters.

In a work involving pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers, Mitra *et al*. [56] obtained several pharmacophore hypotheses having statistically significant values for the R^2_{pred} parameter when mapped with the test set compounds. For the selection of the best hypothesis, the $r_{m\ (test)}^2$ values were calculated and based on the values of this parameter, the final model was selected and reported. In another recent paper by Mitra *et al*. [24], the authors attempted to show that 'true $r_{m\ (LOO)}^2$' statistic calculated based on model derived from undivided data set with the variable selection strategy being applied at each cycle of leave-one-out (LOO) validation reflects external validation characteristics of the developed model thereby eliminating the need of splitting of the data set into training and test sets. A cytoprotection data of anti-HIV thiocarbamates was used as the model data dataset and was divided into 50 different combinations of training and test sets using the *k*-means clustering technique followed by further 25 different combinations by random selection of test set compounds. For each set the value of 'true Q^2' [20] was calculated. 'True Q^2' is a measure of validation developed by Hawkins *et al*. [20] which involves calculation of the Q^2 parameter by applying the variable selection strategy applied at each cycle of leave-one-out (LOO) validation. Based on the activity values thus predicted for each of the training and undivided set compounds, the values of 'true $r_{m\ (LOO)}^2$' were calculated for each of the corresponding sets. Besides these, for each of the combinations, the external predictive parameters were also calculated which include R^2_{pred} and $r_{m\ (test)}^2$. From an analysis of the results obtained, the authors reported that the value of 'true $r_{m\ (LOO)}^2$' (0.425) calculated for the undivided dataset matches with the value of 'mean $r_{m\ (test)}^2$' (0.410) calculated from the 75 different combinations. Moreover, absolute difference of the 'mean $r_{m\ (test)}^2$' value (75 trials) from 'true $r_{m\ (LOO)}^2$' of the model obtained from the undivided data set was insignificant ($p \leq 0.05$). The 'true $r_{m\ (LOO)}^2$' metric thus reflects the criterion for external validation and thereby eliminates the need for splitting of the dataset in training and test sets when the size of the entire dataset is small.

Roy *et al*. [57] performed classical and 3D QSAR studies of cytochrome 17 inhibitor imidazole substituted biphenyls. This work aptly describes that effective splitting of the dataset followed by calculation of the r_m^2 metrics helps in the right selection of the statistically most significant QSAR model. On comparing the results, it is observed that the MFA model, despite bearing a high value for R^2_{pred} parameter (0.876), shows a lower value of the $r_{m\ (overall)}^2$ parameter (0.820). On the contrary, the receptor shape analysis (RSA) model having a maximum Q^2 (0.932) value and a lower value of R^2_{pred} (0.853) parameter exhibits the maximum value for the $r_{m\ (overall)}^2$ (0.858) parameter. This may be explained by the fact that the training set used for modeling being much larger compared to that of the test set, the value of the $r_{m\ (overall)}^2$ parameter is dictated by the LOO predicted activity values of the training set. Thus, although the molecular field analysis (MFA) model bears the maximum value for $r_{m\ (test)}^2$ (0.792) parameter, it experiences a reduced $r_{m\ (overall)}^2$ value due to poor internal predictive power of the model. From these results, it may be inferred that a model which already suffers from poor internal predictivity may be

unsuitable for any future prediction of activity of untested molecules but such a poor predictivity is not reflected in the value of R^2_{pred} parameter. Hence, selection of the best QSAR model based on the value of $r_m^{2}{}_{(overall)}$ parameter reflects both internal and external predictive ability of the developed QSAR model and thus improves its reliability for activity prediction of new molecules. Roy *et al*. [58] performed docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors involving the application of the r_m^2 metrics for the selection of the best QSAR model for this class. The GFA model built using the molecular shape analysis (MSA), spatial, electronic, thermodynamic and structural descriptors yielded a maximum value for Q^2 (0.668), but this model could not achieve the threshold value for the $r_m^{2}{}_{(LOO)}$ (0.496) parameter. This is because the residual activity data for several compounds was more than 1 log unit; $r_m^{2}{}_{(LOO)}$, being solely dependent on this difference between the observed and predicted activity data, efficiently reflects this deviation in its value. Similarly, the GFA model developed using the 2D descriptors exhibited maximum R^2_{pred} (0.687) and $r_m^{2}{}_{(test)}$ (0.657) values, however it suffered from unacceptable $r_m^{2}{}_{(LOO)}$ (0.454). Since the size of the training set is much larger compared to that of the test set, the activity predicted for the training set molecules influences the value of the $r_m^{2}{}_{(overall)}$ parameter more than the test set molecules. Thus, the G/PLS model developed using the MSA, spatial, electronic, thermodynamic and structural descriptors and bearing moderate values for the Q^2 (0.630) and R^2_{pred} (0.630) parameters but the maximum value for the $r_m^{2}{}_{(overall)}$ (0.606) parameter was selected as the best one in terms of both internal and external predictive potential. A maximum value of the $r_m^{2}{}_{(overall)}$ parameter signifies that the observed and the predicted activity data for both the training and test set molecules are located in close proximity to each other.

In another work, Roy *et al*. [59] performed pharmacophore mapping, molecular docking and QSAR studies for a series of structurally diverse compounds as CYP 2B6 inhibitors. For the selection of the best model, the authors calculated the r_m^2 metrics and employed the $r_m^{2}{}_{(overall)}$ parameter for selecting the model with maximum predictive ability. Followed by the generation of pharmacophore hypothesis, QSAR analysis of the molecules was performed using different chemometric tools. In this work, among the three equations obtained, the first one developed using the GFA-spline option yielded the maximum value for the R^2_{pred} (0.843) parameter but a lower value for the $r_m^{2}{}_{(test)}$ (0.676) metric resulting in a reduced value for the $r_m^{2}{}_{(overall)}$ (0.754). Such difference between the values of R^2_{pred} and $r_m^{2}{}_{(test)}$ may be attributed to the fact that a difference of nearly a log unit exists between the observed and predicted activity data for some of the test set molecules (compound nos. **46** and **48**) as determined based on this equation. Thus, it may be inferred that the poor predictive ability of this model is well reflected in the reduced value of $r_m^{2}{}_{(test)}$ parameter. Moreover, the size of the test set being small, the value of the $r_m^{2}{}_{(overall)}$ parameter is influenced to a greater extent by the variation of the predicted activity data of the training set compounds. On the contrary, the second equation developed using the GFA-spline option exhibited the maximum Q^2 (0.772) value together with maximum values for all the three r_m^2 [$r_m^{2}{}_{(test)}$ = 0.749, $r_m^{2}{}_{(LOO)}$ = 0.750, $r_m^{2}{}_{(overall)}$ = 0.774] metrics. Thus it implied that the activity of the compounds predicted using this model closely coincides with the corresponding observed activity data. Hence, despite having a reduced value for the R^2_{pred} (0.832) parameter, the second model was selected as the best one for activity prediction of untested molecules based on the value of $r_m^{2}{}_{(overall)}$ parameter. In yet another paper, Roy *et al*. [60] analysed the statistical quality of the QSAR models based on the calculation of the r_m^2 metrics besides the traditional methods involving the calculation of Q^2 and R^2_{pred} parameters. Besides the docking results, the QSAR models developed for the aromatase inhibitors revealed interesting results. Among all, the third and fourth models developed using the GFA and G/PLS techniques respectively employing 3D and thermodynamic descriptors yielded maximum values for $r_m^{2}{}_{(test)}$ (0.836) and R^2_{pred} (0.864) parameters respectively. But a reduction in the value of $r_m^{2}{}_{(LOO)}$ parameter for both the models (due to extensive deviation between the observed and predicted activity data of the training set molecules) lead to a decline in the values of the $r_m^{2}{}_{(overall)}$ parameter. However, the second model developed using the G/PLS technique employing the 2D and thermodynamic descriptors yielded a maximum value for the $r_m^{2}{}_{(overall)}$ (0.710) metric and thus was selected as the best one in terms of its overall predictive ability.

Kar *et al*. [61] developed a number of QSAR models using the QTMS descriptors calculated at different levels of theory for toxicity prediction of aromatic aldehydes to *Tetrahymena pyriformis*. The developed models were validated using both traditional internal and external validation techniques in addition to calculation of the r_m^2 statistics. In this work, Kar *et al*. developed several comparable models which varied in their predictive ability depending on the values of internal and external predictive parameters (Q^2 and R^2_{pred}). The models with maximum Q^2 values yielded comparatively poorer R^2_{pred} values and vice versa. Thus, the selection of the best model was done based on the value of $r_m^{2}{}_{(overall)}$ parameter. Since this parameter considers the prediction for both the training and test sets, it serves as a more stringent measure for judgment of the predictive potential of the developed QSAR models. Again while developing QSAR models for toxicity of diverse organic chemicals to Daphnia magna using 2D and 3D descriptors, Kar *et al*. [62] developed several QSAR models using different sets of descriptors based on different chemometric tools. Among these, the most predictive QSAR model was selected based on the value of the $r_m^{2}{}_{(overall)}$ metric taking into consideration the predictive ability of both the training and test set molecules.

Another work by Kar *et al*. [63] deals with the first report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. In this work, the authors developed interspecies toxicity correlation between toxicities to Daphnia magna (zooplankton) and fish (species according to OECD guidelines) assessing the ecotoxicological hazard potential of diverse 77 pharmaceuticals. Among the models developed with the Daphnia toxicity, model 7 developed using the G/PLS linear technique was selected as the best one in terms of its acceptable values for all the internal and external validation parameters. A comparison of models 6

and 7 for Daphnia toxicity revealed that model 6, despite bearing maximum value for Q^2 (0.707) exhibited unacceptable value for the $r_{m(overall)}^2$ (0.484) parameter. A difference of approximately 1 log unit between the observed and predicted activity data for some of the test set compounds accounts for such unacceptable values of the r_m^2 metrics for model 6. On the contrary, in case of model 7, a little difference in the values of each of the internal (Q^2) and external (R^2_{pred}) predictive parameters with the respective r_m^2 metric signifies that the predicted and the observed activity data for both the whole training and test sets are located in close vicinity to each other and hence the model shows the maximum value for the $r_{m(overall)}^2$ (0.543) parameter. Similarly, amongst the models developed using the fish toxicity data as the dependent variable, model 12 developed using the G/PLS linear technique exhibited maximum values for all the internal and external validation parameters as well as for all the three r_m^2 metrics and hence it was selected as the most predictive one.

Dashtbozorgi *et al*. [64] predicted air to liver partition coefficient for volatile organic compounds using QSAR approaches based on two different chemometric tools, partial least squares (PLS) and artificial neural network (ANN), followed by validation employing different internal and external predictive parameters. Besides the different external validation techniques of Golbraikh and Tropsha [21], they also utilized the $r_{m(test)}^2$ (0.980) metric for selection of model with maximum external predictivity. The ANN models with a large value of the $r_{m(test)}^2$ parameter was thus selected since ANN model could describe more accurately the relationship between the structural parameters and air to liver partition coefficients of volatile organic compounds. Moreover, an acceptable value of the $r_{m(test)}^2$ parameter indicates that the difference between the observed and predicted activity data of the test set compounds is minimum. In another report, Golmohammadi *et al*. [65] studied the QSAR prediction of gas-to-chloroform partition coefficient using artificial neural network. They employed a PLS method for the selection the best descriptors, which were used as input neurons in neural network model, and selected the final model based on different external predictive parameters which included the calculation of the $r_{m(test)}^2$ metric. An ANN study performed with these molecules yielded the final best QSAR model with a significant value of $r_{m(test)}^2$ metric equal to 0.924. The result indicates a close proximity between the predicted and observed activity values of the test set compounds. Different other statistical parameters confirm superiority of the ANN model over the PLS model. In yet another work, Golmohammadi *et al*. [66] assessed the prediction of inherent viscosity for polymers containing natural amino acids from the theoretically derived molecular descriptors and reported the final best QSAR model based on the calculation different validation parameters including $r_{m(test)}^2$. In this work, the descriptors chosen by genetic algorithm (GA) and multiple linear regression (MLR) feature selection techniques were used as inputs for the subsequent neural network. The model developed using the ANN technique yielded a statistically significant value of $r_{m(test)}^2$ metric (0.921) revealing superiority of the ANN model over the linear model. Arkan *et al*. [67] reported validated QSAR analysis of some diaryl substituted pyrazoles as CCR2 inhibitors by various linear and nonlinear multivariate chemometrics methods. For the selection of the best model, the authors validated the models employing different validation parameters. They utilized different parameters as reported by Tropsha [21] and Roy [22] for assessing the predictive ability of the developed QSAR models. Amongst the different linear and non-linear models developed, the model developed using the least squares support vector machine (LS-SVM) technique yielded the best results in terms of R^2 for both the training (0.911) and test sets (0.861) and acceptable values for the r_m^2 metrics [$r_{m(LOO)}^2$ = 0.638 and $r_{m(test)}^2$ = 0.564]. The results thus show that the observed and predicted activity data for the test set compounds closely coincide with each other indicating superior predictivity of LS-SVM models over the other models developed in the work. Thus, other models, despite showing acceptable values of the internal predictive parameters, suffer from poor predictivity as reflected by the values of the modified r^2 (r_m^2). Hence, it may be stated that for closely related models showing comparable values for the different statistical parameters, the r_m^2 metric serves as a crucial measure for the selection of the most significant QSAR model.

Ojha *et al*. [68], in their work involving the chemometric modeling, docking and *in silico* design of triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors as antimalarials, selected the final best developed QSAR model based on the value of $r_{m(overall)}^2$ preferably over that of R^2_{pred}. Among the two classical QSAR models developed using the G/PLS spline option, the first one (bearing the MR_p, MR_m, $B1_p$, π_m, L_o descriptors) exhibited maximum value for $r_{m(overall)}^2$ (0.733) and was selected as the best developed model, though the R^2_{pred} value for the model (0.767) was much lower compared to the next one (0.824). Such an observation may be attributed to the fact that in case of the second model (developed using the $B1_p$, $B5_o$, MR_p, $B1_m$, π_p descriptors), a significant difference exists between the values of R^2_{pred} (0.824) and $r_{m(test)}^2$ (0.788) parameters indicating a considerable deviation between the values of the predicted and the observed activity for the test set molecules.

Instead of R^2_{pred}, $r_{m(test)}^2$ has been used by Khosrokhavar *et al*. [69] as the metric for selection of most predictive QSAR models in their 2D QSAR study for mycotoxins using multiple linear regression and support vector machine. Both the models achieved acceptable values for all the parameters reported by Golbraikh and Tropsha [21] as well as the three r_m^2 metrics. However, among the two models, the model developed using the regression technique is more predictive compared to the support vector machine (SVM) model. This can be inferred from the observation that in case of the MLR model, a few compounds exhibit a difference of more than 2 log units between the observed and predicted activity data, while in case of the SVM model such a difference exists in a comparatively greater number of compounds. Hence the value of $r_{m(test)}^2$ (which solely depends on the observed and predicted activity data of the test set molecules) is more for the MLR (0.894) model than that for the SVM (0.833) model. Goodarzi *et al*. [70] developed PLS and N-PLS (multilinear PLS) based MIA-QSTR (multivariate image analysis-quantitative structure-toxicity relationship) modeling of the acute toxicities of phenylsulphonyl

carboxylates to *Vibrio fischeri* and employed the r_m^2 metric for assessing the predictive potential of the developed models. The results revealed that the r_m^2 values for both the N-PLS and PLS models were similar, both for the training and test sets: 0.742 and 0.758 for the training set (N-PLS and PLS, respectively), and 0.584 and 0.581 for the test set (N-PLS and PLS, respectively). The authors thus concluded that both the models were equally predictive in terms of the r_m^2 validation parameters. Lan *et al*. [71] analysed the r_m^2 metric for the development of molecular models using *d*-annulated benzazepinones as vascular endothelial growth factor receptor 2 (VEGF-R2) kinase inhibitors based on 3D-QSAR techniques. From the results, the authors inferred that both the CoMFA and CoMSIA models were well predictive in terms of their external predictive parameter (R^2_{pred}) but exhibited distinct variation when the $r_m^2{}_{(test)}$ parameter was calculated. A higher value of the $r_m^2{}_{(test)}$ parameter for the CoMFA model signified that the activity of the test set molecules predicted using the CoMFA model closely matched with the observed data. On the contrary, large residual values for the activity data predicted using the CoMSIA model may be attributed to the reduced potential of the CoMSIA model for activity prediction of untested molecules.

In a more recent work, Pran Kishore *et al*. [72] developed QSAR models using MFA and MSA techniques for adenosine receptor antagonists exploring physicochemical requirements for binding of pyrazolo[4,3-e]-1,2,4-triazolo[1,5-c] pyrimidine derivatives with human adenosine A3 receptor subtype. Various validation measures were performed for the selection of the statistically significant QSAR models. The authors reported that the MSA model was the best one in terms of external predictive ability (R^2_{pred} = 0.760) while the MFA model was the most significant one in terms of internal predictive power (Q^2 = 0.762). Although large values were obtained for the internal predictive parameter (Q^2), comparatively reduced values for the $r_m^2{}_{(LOO)}$ metric reflected that a considerable difference existed between the observed and LOO predicted activity data of the training set compounds. However, in terms of overall predictive ability [$r_m^2{}_{(overall)}$ = 0.593], the MFA model was reported to be the best one.

Besides these, several other studies reported by different authors [73-138] reveal the importance of the r_m^2 metrics for the selection of the best QSAR models. The r_m^2 parameters exhibit a direct comparison between the observed and predicted activity data of the training and test set compounds and these parameters offer more stringent tests than the classical validation parameters especially for data sets with wide ranges of response values [23]. Hence, these parameters have been used for model selection by several groups of researchers. Very recently [139], additional variants of r_m^2 parameters have been proposed; however, these have not been covered in the present review.

CONCLUSION

QSAR models have been traditionally tested for their predictive potential using internal (Q^2) and external validation (R^2_{pred}) parameters. These parameters bear extensive information regarding the ability of a QSAR model to predict the activity of untested molecules. But being primarily dependent on the mean activity data of the training set compounds, both the parameters tend to automatically achieve acceptable values (> 0.5) whenever a data set with a wide range of activity data is considered. However, such values may not truly reflect the extent of deviation of the predicted activity values from the observed ones. Therein lies the utility of the r_m^2 metrics. It has been observed that the traditional internal and external validation parameters exhibit acceptable values as long as an overall good correlation is maintained between the observed and the predicted activity data irrespective of the actual difference between the two values. However, the r_m^2 metrics depend chiefly on the difference between the observed and predicted activity data and convey more accurate information regarding the deviation between the two values. The $r_m^2{}_{(LOO)}$ parameter compares between the observed activity data of the training set compounds and their LOO predicted activity, thereby signifying the internal predictive ability of the developed QSAR model. On the other hand, the $r_m^2{}_{(test)}$ metric determines the proximity between the values of the observed and predicted activity of the test set compounds. Thus, the inability of the Q^2 and R^2_{pred} metrics to reflect bad predictions for some compounds suggests that the parameters $r_m^2{}_{(LOO)}$ and $r_m^2{}_{(test)}$ are stricter metrics for validation in comparison to the traditional ones. Moreover, the $r_m^2{}_{(overall)}$ metric is an unique parameter considering predictions for both training and test set compounds and its value is not obtained from prediction of limited number of test set compounds as is the case for R^2_{pred}. Having the ability to reflect the predictive ability of the model in terms of both internal and external validation, the $r_m^2{}_{(overall)}$ parameter can be aptly utilized to identify the best QSAR model from among comparable models, especially when different models show different patterns in internal and external predictivity. In the present review, it has been reported that in several studies by different authors, the r_m^2 metrics have been utilized for the selection of the final QSAR models. Different authors have reported that several models bear unacceptable values of the $r_m^2{}_{(test)}$ parameter despite having statistically significant values for R^2_{pred}. Moreover in several cases, comparable models were obtained with different patterns in Q^2 and R^2_{pred} values. Thus, in such case, the $r_m^2{}_{(overall)}$ parameter plays a crucial role for selecting the best model, well predictive in terms of both internal and external predictive ability. Consequently, it may be inferred that this new set of r_m^2 parameters serve as a stricter metric for assessing the predictive potential of the developed QSAR models. Thus, in addition to the traditional validation parameters, tests for the r_m^2 metrics should be carried out for a more stringent test of validation of predictive QSAR models, especially when a regulatory decision is involved.

DECLARATION OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENT

The authors thank Indian Council of Medical Research (ICMR), New Delhi for a senior research fellowship to IM.

REFERENCES

[1] Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, **1962**, *194,* 178-180.

[2] Todeschini, R.; Consonni, V. In: *Handbook of Molecular Descriptors*, Mannhold, R., Kubinyi, H., Timmerman, H., Ed.; Wiley-VCH: Weinheim, **2000**.

[3] Helguera, A.M.; Combes, R.D.; Gonzalez, M.P.; Cordeiro, M.N. Applications of 2D descriptors in drug design: a DRAGON tale. *Curr. Top. Med. Chem*., **2008**, *8*, 1628–1655

[4] Gonzalez, M.P.; Teran, C.; Saiz-Urra, L.; Teijeira, M. Variable selection methods in QSAR: an overview. *Curr. Top. Med. Chem*., **2008**, *8*, 1606–1627.

[5] Perkins, R.; Fang, H.; Tong, W.; Welsh, W.J. Quantitative structure-activity relationship methods: perspectives on drug discovery and toxicology. *Environ. Toxicol. Chem.*, **2003,** *22*, 1666-1679.

[6] Yang, G.F.; Huang, F. Development of Quantitative Structure-Activity Relationships and Its Application in Rational Drug Design. *Curr. Pharm. Des.*, **2006**, *12*, 4601-4611.

[7] Mazzatorta, P.; Benfenati, E.; Lorenzini, P.; Vighi, M. QSAR in ecotoxicity: an overview of modern classification techniques. *J. Chem. Inf. Comput. Sci.*, **2004**, *44*, 105-112.

[8] Walker, J.D.; Dearden, J.C.; Schultz, T.W.; Jaworska, J.; Comber, M.H.I. In: *QSARs for Pollution Prevention, Toxicity Screening, Risk Assessment, and Web Applications*, Walker, J.D., Ed.; SETAC Press: Pensacola, FL, **2003**, pp. 3–18

[9] Walker, J.D.; Jaworska, J.; Comber, M.H.I.; Schultz, T.W.; Dearden, J.C. Guidelines for developing and using quantitative structure–activity relationships. *Environ. Toxicol. Chem.*, **2003**, *22*, 1653–1665.

[10] Cronin, M.T.D.; Schultz, T.W. Pitfalls in QSAR. *J. Theoret. Chem. (Theochem)*, **2003**, 622, 39–51.

[11] Eriksson, L.; Jaworska, J.; Worth, A.P.; Cronin, M.T.D.; McDowell, R.M.; Gramatica, P. Methods for reliability and uncertainty assessment for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Persp.*, **2003**, *111*, 1361–1375.

[12] Livingstone, D.J. In: *Building QSAR models: a practical guide, in Predicting Chemical Toxicity and Fate*, Cronin, M.T.D., Livingstone, D.J., Eds.; CRC Press: Boca Raton, FL, **2004**; pp. 151–170.

[13] Tropsha, A.; Golbraikh, A. Predictive QSAR modeling workflow, model applicability domains, and virtual screening. *Curr. Pharmaceutical Design*, **2007**, *13*, 3494–3504.

[14] Wold, S.; Eriksson, L. In: *Chemometrics Methods in Molecular Design*, van de Waterbeemd, H., Ed.; VCH: Weinheim, Germany, **1995**; pp. 309-318.

[15] Zvinavashe, E.; Murk, A.J.; Rietjens, I.M.C.M. Promises and pitfalls of quantitative structureactivity relationship approaches for predicting metabolism and toxicity. *Chem. Res. Toxicol.*, **2008**, *21*, 2229-2236.

[16] Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.*, **2007**, *26*, 694-701.

[17] http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm (Accessed Nov 29, 2010)

[18] OECD Principles for the Validation of (Q)SARs, http://www.oecd.org/dataoecd/33/37/37849783.pdf (Accessed Nov 29, 2010)

[19] Roy, K. On some aspects of validation of predictive QSAR models. *Expert Opin. Drug Discov.,* **2007**, *2*, 1567-1577.

[20] Hawkins, D.M.; Basak, S.C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.*, **2003**, 43, 579–586.

[21] Golbraikh, A.; Tropsha, A. Beware of q^2! *J. Mol. Graphics Mod.*, **2002**, *20*, 269-276.

[22] Roy, P.P; Roy, K. On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.*, **2008**, *27*,302–313

[23] Roy, P.P.; Paul, S.; Mitra, I.; Roy, K. On two novel parameters for validation of predictive QSAR models. *Molecules*, **2009**, *14,*1660-1701.

[24] Mitra, I.; Roy, P.P.; Kar, S.; Ojha, P.; Roy, K. On further application of r_m^2 as a metric for validation of QSAR models. *J. Chemometrics*, **2010,** *24***,** 22-33.

[25] Topliss, J.G.; Costello, R.J. Chance correlation in structure-activity studies using multiple regression analysis. *J. Med. Chem.*, **1972,** *15*(10), 1066-1068

[26] Golbraikh, A.; Shen, M.; Xiao, Z.; Xiao, Y.D.; Lee, K.H.; Tropsha., A. J. Rational selection of training and test sets for the development of validated QSAR models. *Comput. Aided Mol. Design*, **2003**, *17*, 241 – 253.

[27] Zheng, W.; Tropsha, A. A. novel variable selection quantitative structureproperty relationship approach based on the *k*-nearest-neighbor principle. *J. Chem. Inf. Comput. Sci.*, **2000**, *40*, 185 – 194.

[28] Ghafourian, T.; Cronin, M. T. The impact of variable selection on the modelling of oestrogenicity. *SAR QSAR Environ. Res.*, **2005**, *16*, 171 – 190.

[29] He, L.; Jurs, P. C. Assessing the reliability of a QSAR model's predictions. *J. Mol. Graphics Model.*, **2005**, *23*, 503 –523.

[30] Guha, R.; Jurs, P.C. Determining the validity of a QSAR model-a classification approach. *J. Chem. Inf. Model.*, **2005**, *45*, 65 – 73.

[31] Aptula, A.O.; Jeliazkova, N.G.; Schultz, T.W.; Cronin, M. T. D. The better predictive model: High Q^2 for the training set or low root mean square error of prediction for the test set? *QSAR Comb. Sci.*, **2005**, *24*, 385 – 396.

[32] Gramatica, P.; Giani, E.; Papa, E. Statistical external validation and consensus modeling: a QSPR case study for Koc prediction. *J. Mol. Graphics. Model.*, **2007**, *25*, 755 – 766.

[33] Nicholls, A.; MacCuish, N. E.; MacCuish, J. D. Variable selection and model validation of 2D and 3D molecular descriptors. J. *Comput. Aided Mol. Design*, **2004**, *18*, 451 – 474.

[34] Shen, M.; Beguin, C.; Golbraikh, A.; Stables, J. P.; Kohn, H.; Tropsha, A. Application of predictive QSAR models to **...** identification and experimental validation of novel anticonvulsant compounds. *J. Med. Chem.*, **2004**, *47*, 2356 – 2364.

[35] Debnath, A.K. In: *Combinatorial Library Design and Evaluation*, Ghose A.K., Viswanadhan V.N., Eds.; Marcel Dekker: New York, **2001**

[36] Tong, W.; Xie, Q.; Hong, H.; Shi, L.; Fang, H.; Perkins, R. Assessment of prediction confidence and domain extrapolation of two structure–activity relationship models for predicting estrogen receptor binding activity. *Environ. Health Perspect.*, **2004**, *112*, 1249 -1254

[37] Tropsha, A.; Gramatica, P.; Gombar, V. K. The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci*., **2003**, *22*(1), 69–77

[38] Roy, P.P.; Leonard, J.T.; Roy, K. Exploring the impact of the size of training sets for the development of predictive QSAR models. *Chemom. Intell. Lab. Syst.*, **2008**, *90*, 31–42.

[39] Roy, K.; Roy, P.P. Comparative QSAR studies of CYP1A2 inhibitor flavonoids using 2D and 3D descriptors. *Chem. Biol. Drug Des.*, **2008,** *72*, 370–382.

[40] Roy, K.; Popelier, P.L.A. Exploring predictive QSAR models for hepatocyte toxicity of phenols using QTMS descriptors. *Bioorg. Med. Chem. Lett.*, **2008**, *18*, 2604–2609.

[41] Roy, K.; Popelier, P.L.A. Exploring predictive QSAR models using quantum topological molecular similarity (QTMS) descriptors for toxicity of nitroaromatics to *Saccharomyces cerevisiae*. *QSAR Comb. Sci.*, **2008**, *27*, 1006 – 1012.

[42] Roy, K.; Popelier, P.L.A. Predictive QSPR modeling of the acidic dissociation constant (pKa) of phenols in different solvents. *J. Phys. Org. Chem.*, **2009**, *22*, 186–196.

[43] Roy K.; Paul; S. Exploring 2D and 3D QSARs of 2, 4-Diphenyl-1,3-oxazolines for Ovicidal Activity Against *Tetranychus urticae*. *QSAR Comb. Sci.*, **2009**, *28*, 406 – 425.

[44] Basu, A.; Jasu, K.; Jayaprakash, V.; Mishra, N.; Ojha, P.; Bhattacharya, S. Development of CoMFA and CoMSIA models of cytotoxicity data of anti-HIV-1-phenylamino-1H-imidazole derivatives. *Eur. J. Med. Chem.*, **2009,** 44, 2400–2407.

[45] Roy, K.; Roy, P.P. Comparative chemometric modeling of cytochrome 3A4 inhibitory activity of structurally diverse compounds using stepwise MLR, FA-MLR, PLS, GFA, G/PLS and ANN techniques. *Eur. J. Med. Chem.*, **2009,** *44*, 2913–2922.

[46] Roy, K.; Roy, P.P. Exploring QSAR and QAAR for inhibitors of cytochrome P450 2A6 and 2A5 enzymes using GFA and G/PLS techniques. *Eur. J. Med. Chem.*, **2009**, *44*, 1941–1951.

[47] Hu, R.; Doucet, J-P.; Delamar, M.; Zhang, R. QSAR models for 2-amino-6-arylsulfonylbenzonitriles and congeners HIV-1 reverse transcriptase inhibitors based on linear and nonlinear regression methods. *Eur. J. Med. Chem.*, **2009**, *44,* 2158–2171

[48] Mitra, I.; Saha A.; Roy, K. Quantitative Structure–Activity Relationship Modeling of Antioxidant Activities of Hydroxybenzalacetones Using Quantum Chemical, Physicochemical and Spatial Descriptors. *Chem. Biol. Drug Des.*, **2009,** *73,* 526–536

[49] Mitra, I.; Saha A.; Roy, K. QSAR of antilipid peroxidative activity of substituted benzodioxoles using chemometric tools. *J. Comput. Chem.*, **2009,** *30*, 2712–2722.

[50] Srivastava, M.; Singh H.; Naik, P.K. Quantitative structure–activity relationship (QSAR) of artemisinin: the development of predictive *in vivo* antimalarial activity models. *J. Chemometrics* , **2009**, *23*, 618–635.

[51] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSPR modeling bioconcentration factor (BCF) by balance of correlations. *Eur. J. Med. Chem.*, **2009**, *44*, 2544–2551

[52] Nargotra, A.; Sharma, S.; Koul, J.L.; Sangwan, P.L.; Khan, I.A.; Kumar, A.; Taneja, S.C.; Koul, S. Quantitative structure activity relationship (QSAR) of piperine analogs for bacterial NorA efflux pump inhibitors. *Eur. J. Med. Chem.*, **2009**, *44*, 4128–4135

[53] Nargotra, A.; Koul, S.; Sharma, S.; Khan, I.A.; Kumar, A.; Thota, N.; Koul, J.L.; Taneja, S.C.; Qazi, G.N. Quantitative structureeactivity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors. *Eur. J. Med. Chem.*, **2009**, *44*, 229-238.

[54] Roy K.; Paul; S. Docking and 3D QSAR studies of protoporphyrinogen oxidase inhibitor 3H-pyrazolo [3,4-d][1,2,3]triazin-4-one derivatives. *J. Mol. Model.*, **2010**, *16*, 137–153

[55] Liao, S.Y.; Chen , J.C.; Miao, T.F.; Shen, Y.; Zheng, K.C. Binding conformations and QSAR of CA-4 analogs as tubulin inhibitors. *J. Enzyme Inhib. Med. Chem.*, **2010**, *25*, 421-429.

[56] Mitra, I.; Saha A.; Roy, K. Pharmacophore mapping of arylamino-substituted benzo[b]thiophenes as free radical scavengers. *J. Mol. Model.*, **2010**, *16*, 1585–1596 .

[57] Roy, P.P.; Roy, K. Classical and 3D-QSAR studies of cytochrome 17 inhibitor imidazole-substituted biphenyls. *Mol. Simul.,* **2010**, 36, 311–325.

[58] Roy, P.P.; Roy, K. Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *J. Mol. Model.*, **2010**, *16*, 1597-1616.

[59] Roy, P.P.; Roy, K. Pharmacophore mapping, molecular docking and QSAR studies of structurally diverse compounds as CYP2B6 inhibitors. *Mol. Simul.,* **2010**, *36*, 887-905

[60] Roy, P.P.; Roy, K. Molecular docking and QSAR studies of aromatase inhibitor androstenedione derivatives. *J. Pharm. Pharmacol.*, **2010**, 62, 1717-1728.

[61] Kar, S.; Harding, A.P.; Roy K.; Popelier, P.L.A. QSAR with quantum topological molecular similarity indices: toxicity of aromatic aldehydes to Tetrahymena pyriformis; *SAR QSAR Environ. Res.*, **2010**, *21*, 149 – 168.

[62] Kar, S.; Roy, K. QSAR modeling of toxicity of diverse organic chemicals to Daphnia magna using 2D and 3D descriptors. *J. Hazard. Mater.*, **2010**, *177* 344–351

[63] Kar, S.; Roy; K. First report on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. *Chemosphere*, **2010**, *81*, 738-747.

[64] Dashtbozorgi, Z.; Golmohammadi, H. Prediction of air to liver partition coefficient for volatile organic compounds using QSAR approaches. *Eur. J. Med. Chem.*, **2010,** 45, 2182–2190.

[65] Golmohammadi, H.; Safdari, M. Quantitative structure–property relationship prediction of gas-to-chloroform partition coefficient using artificial neural network. *Microchem. J*., **2010**, *95*, 140–151.

[66] Mallakpour, S.; Hatami, M.; Golmohammadi, H. Prediction of inherent viscosity for polymers containing natural amino acids from the theoretical derived molecular descriptors. *Polymer*, **2010**, *51*, 3568-3574.

[67] Arkan, E.; Shahlaei, M.; Pourhossein, A.; Fakhri, K.; Fassihi, A. Validated QSAR analysis of some diaryl substituted pyrazoles as CCR2 inhibitors by various linear and nonlinear multivariate chemometrics methods. *Eur. J. Med. Chem.*, **2010**, *45*, 3394-3406

[68] Ojha, P.K.; Roy, K.Chemometric modeling, docking and in silico design of triazolopyrimidine-based dihydroorotate dehydrogenase inhibitors as antimalarials. *Eur. J. Med. Chem.*, **2010**, *45*, 4645-4656.

[69] Khosrokhavar, R.; Ghasemi, J.B.; Shiri, F. 2D Quantitative Structure-Property Relationship Study of Mycotoxins by Multiple Linear Regression and Support Vector Machine. *Int. J. Mol. Sci.*, **2010**, *11*, 3052-3068.

[70] Goodarzi, M.; Freitas, M. P. PLS and N-PLS-based MIA-QSTR modelling of the acute toxicities of phenylsulphonyl carboxylates to Vibrio fischeri. *Mol. Simul.*, **2010**, *36*, 953-959.

[71] Lan, P.; Sun, J-R.; Chen, W-N.; Sun, P-H.; Chen, W-M. Molecular modelling studies on d-annulated benzazepinones as VEGF-R2 kinase inhibitors using docking and 3D-QSAR. *J. Enzyme Inhib. Med. Chem.*, **2011**, *26*, 367-377.

[72] Prankishore, D.; Balakumar, C.; Rao, A.R.; Roy, P.P.; Roy, K. QSAR of adenosine receptor antagonists: Exploring physicochemical requirements for binding of pyrazolo[4,3-*e*]-1,2,4-triazolo[1,5-*c*]pyrimidine derivatives with human adenosine A3 receptor subtype. *Bioorg. Med. Chem. Lett.*, **2010**, *21*, 818-823.

[73] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSPR modeling for enthalpies of formation of organometallic compoundsby means of SMILES-based optimal descriptors. *Chem. Phys. Lett.*, **2008**; *461,* 343–347.

[74] Lagos, C.F.; Caballero, J.; Gonzalez-Nilo, F.D.; David Pessoa-Mahana, C.; Perez-Acle, T. Docking and Quantitative Structure–Activity Relationship Studies for the Bisphenylbenzimidazole Family of Non-Nucleoside Inhibitors of HIV-1 Reverse Transcriptase. *Chem. Biol. Drug. Des*., **2008**, *72*, 360–369.

[75] Goodarzi, M.; Puggina de Freitas, M. MIA-QSAR modelling of activities of a series of AZT analogues: bi- and multilinear PLS regression. *Mol. Simul.*, **2010**, *36*, 267–272.

[76] Nargotra, A.; Koul, S.; Sharma, S.; Khan, I.A.; Kumar, A.; Thota, N.; Koul, J.L.; Taneja, S.C.;, Qazi, G.N. Quantitative structure eactivity relationship (QSAR) of aryl alkenyl amides/imines for bacterial efflux pump inhibitors. *Eur. J. Med. Chem.,* **2009**, *44*, 229- 238.

[77] Naik, P.K.; Sidhura; Singh, T.; Singh, H. Quantitative structure–activity relationship (QSAR) for insecticides: development of predictive *in vivo* insecticide activity models. *SAR QSAR Environ. Res.*, **2009**, *20*, 551-566. .

[78] Zeng, H.; Zhang, H. Combined 3D-QSAR modeling and molecular docking study on 1,4-dihydroindeno[1,2-c]pyrazoles as VEGFR-2 kinase inhibitors. *J. Mol. Graph. Model.*, **2010**, *29*, 54-71.

[79] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. SMILES-based optimal descriptors: QSAR analysis of fullerene-based HIV-1 PR inhibitors by means of balance of correlations. *J. Comput. Chem.*, **2010,** *31*, 381-392.

[80] Toropov, A.A.; Toropova, A.P.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. InChI-based optimal descriptors: QSAR analysis of fullerene[C60]-based HIV-1 PR inhibitors by correlation balance. *Eur. J. Med. Chem*., **2010**, *45,* 1387-1394.

[81] Lu, P.; Wei, X.; Zhang, R. CoMFA and CoMSIA 3D-QSAR studies on quionolone caroxylic acid derivatives inhibitors of HIV-1 integrase. *Eur. J. Med. Chem.,* **2010**, *45*, 3413-3419. Sharma, B.S.; Pilania, P.; Sarbhai, K.; Singh, P.; Prabhakar, Y.S. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. *Mol. Divers.*, **2009,** *14*, 371-384.

[82] Cheng, J.; Liu, G.; Zhang, J.; Xu, Z.; Tang, Y. Insights into subtype selectivity of opioid agonists by ligand-based and structure-based methods. *J. Mol. Model.*, **2011**, *17*, 477–493.

[83] Toropov, A.A.; Toropova, A.P.; Benfenati, E. QSPR modelling of normal boiling points and octanol/water partition coefficient for acyclic and cyclic hydrocarbons using SMILES-based optimal descriptors. *Cent. Eur. J. Chem.*, **2010**, *8*, 1047–1052 .

[84] Tromelin, A.; Merabtine, Y.; Andriot, I.; Lubbers, S.; Guichard, E. Retention–release equilibrium of aroma compounds in polysaccharide gels: study by quantitative structure–activity/property relationships approach. *Flavour Fragr. J.,* **2010**, *25*, 431-442.

[85] Toropova, A.P.; Toropov, A.A.; Lombardo, A.; Roncaglioni, A.; Benfenati, E.; Gini, G. A new bioconcentration factor model based

on SMILES and indices of presence of atoms. *Eur. J. Med. Chem.*, **2010**, *5*, 4399-402.

[86] Shahlaei, M.; Sabet, R.; Ziari, M.B.; Moeinifard, B.; Fassihi, A.; Karbakhsh, R. QSAR study of anthranilic acid sulfonamides as inhibitors of methionine aminopeptidase-2 using LS-SVM and GRNN based on principal components. *Eur. J. .Med. Chem.*, **2010**, *45*, 4499-4508.

[87] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. QSAR modeling of measured binding affinity for fullerene-based HIV-1 PR inhibitors by CORAL. *J. Math. Chem.,* **2010**, 48, 959–987.

[88] Lan, P.; Chen, W-N.; Huang, Z-J.; Sun, P-H.; Chen, W-M .Understanding the structure-activity relationship of betulinic acid derivatives as anti-HIV-1 agents by using 3D-QSAR and docking. *J. Mol. Model.*, **2011**, *17*, 1643-1659.

[89] Deeb, O.; Goodarzi, M. Predicting the solubility of pesticide compounds in water using QSPR methods. *Mol. Phys.*, **2010**, *108*, 181–192.

[90] Halder, A.K.; Jha, T. Validated predictive QSAR modeling of N-aryl-oxazolidinone-5-carboxamides for anti-HIV protease activity. *Bioorg. Med. Chem. Lett.*, **2010**, *20*, 6082-6087.

[91] Ravichandran, V.; Mourya, V.K.; Agrawal, R.K. Prediction of HIV-1 protease inhibitory activity of 4-hydroxy-5,6-dihydropyran-2-ones: QSAR study. *J. Enz. Inhib. Med. Chem*., **2011**, *26*, 288-294.

[92] Kumar, S.; Singh, V.; Tiwari, M. QSAR modeling of the inhibition of reverse transcriptase enzyme with benzimidazolone analogs. *Med. Chem. Res.*, **2011**, *20*, 1530-1541.

[93] Roy, K.; Ghosh, G. QSTR with extended topochemical atom indices 10. Modeling of toxicity of organic chemicals to humans using different chemometric tools. *Chem. Biol. Drug Des.*, **2008,** *72,* 383-394.

[94] Roy, K.; Mitra, I.; Saha, A. Molecular shape analysis of antioxidant and squalene synthase inhibitory activities of aromatic tetrahydro-1,4-oxazine derivatives. *Chem. Biol. Drug Des.,* **2009**, *74*, 507-516.

[95] Ray, S.; Roy, P.P.; Sengupta, C.; Roy, K. Exploring QSAR of hydroxyphenylureas as antioxidants using physicochemical and electrotopological state (E-State) atom parameters. *Mol. Simul.***, 2010,** *36*, 484–492**,**

[96] Ojha, P.K.; Roy, K. Chemometric modelling of antimalarial activity of aryltriazolylhydroxamates. *Mol. Simul.,* **2010,** *36*, 939-952.

[97] Mitra, I.; Saha, A.; Roy, K. Exploring quantitative structure-activity relationship (QSAR) studies of antioxidant phenolic compounds obtained from traditional Chinese medicinal plants. *Mol. Simul.,* **2010,** *36*, 1067–1079**.**

[98] Roy, K.; Das, R.N. QSTR with extended topochemical atom (ETA) indices. 14. QSAR modeling of toxicity of aromatic aldehydes to *Tetrahymena pyriformis*. *J. Hazard. Mater.* **2010,** *183*, 913-922.

[99] Mitra, I.; Saha, A.; Roy, K**.;** Chemometric modeling of free radical scavenging activity of flavone derivatives. *Eur. J. Med. Chem.***, 2010,** *45*, 5071-5079.

[100] Roy, P.P.; Roy, K. QSAR Studies of CYP2D6 Inhibitor Aryloxypropanolamines Using 2D and 3D Descriptors. *Chem. Biol. Drug Des.*, **2009**, *73*, 442–455.

[101] Roy, P.P.; Roy, K. Exploring QSAR for CYP11B2 binding affinity and CYP11B2/CYP11B1 selectivity of diverse functional compounds using GFA and G/PLS techniques. *J. Enzyme Inhib. Med. Chem.*, **2010**, *25*, 354–369.

[102] Mandal, A.S.; Roy, K. Predictive QSAR modeling of HIV reverse transcriptase inhibitor TIBO derivatives. *Eur. J. Med. Chem.*, **2009**, *44*, 1509-1524.

[103] Paul, S.; Roy, K. Docking and 3D-QSAR studies of acetohydroxy acid synthase inhibitor sulfonylurea derivatives. *J. Mol. Model.*, **2010,** *16*, 951-964.

[104] Garriga, M.; Caballero, J. Insights into the structure of urea-like compounds as inhibitors of the juvenile hormone epoxide hydrolase (JHEH) of the tobacco hornworm *Manduca sexta*: Analysis of the binding modes and structure–activity relationships of the inhibitors by docking and CoMFA calculations. *Chemopshere*, **2011**, *82*, 1604-1613.

[105] Deeb, O.; Khadikar, P.V.; Goodarzi, M. QSPR Modeling of Bioconcentration Factors of Nonionic Organic Compounds. *Environ. Health Insights*, **2010**, *4*, 33–47.

[106] Lan, P.; Chen, W-N.; Chen, W-M. Molecular modeling studies on imidazo [4,5-b] pyridine derivatives as Aurora Akinase inhibitors using 3D-QSAR and docking approaches. *Eur. J. Med. Chem.*, **2011**, *46*, 77-94,

[107] Alavi, A.H.; Ameri, M.A.; Gandomi, H.; Mirzahosseini, M.R. Formulation of flow number of asphalt mixes using a hybrid computational method. *Constr. Build. Mater.*, **2011,** *25*, 1338–1355.

[108] Saghaie, L.; Shahlaei, M.; Fassihi, A.; Madadkar- Sobhani, A.; Gholivand, M.B.; Pourhossein, A. QSAR analysis for some diaryl-substituted pyrazoles as CCR2 Inhibitors by GA-Stepwise MLR. *Chem. Biol. Drug Des.*, **2011,** *77*, 75–85

[109] Dashtbozorgi, Z.; Golmohammadi, H. Quantitative structure–property relationship modeling of water-to-wet butyl acetate partition coefficient of 76 organic solutes using multiple linear regression and artificial neural network. *J. Sep. Sci.*, **2010**, *33*, 3800–3810.

[110] Saghaie, L.; Shahlaei, M.; Madadkar-Sobhani, A.; Fassihi, A. Application of partial least squares and radial basis function neural networks in multivariate imaging analysis-quantitative structure activity relationship: Study of cyclin dependent kinase 4 inhibitors. *J. Mol. Graph. Model.*, **2010**, *29*, 518–528.

[111] Golmohammadi, H.; Dashtbozorgi, Z. Quantitative structure–property relationship studies of gas-to-wet butyl acetate partition coefficient of some organic compounds using genetic algorithm and artificial neural network. *Struct. Chem.*, **2010**, *21*, 1241–1252.

[112] Ai, Y.; Song, F-J.; Wang, S-T.; Sun, Q.; Sun, P-H. Molecular modeling studies on 11*H*-dibenz[*b*,*e*]azepine and dibenz[*b*,*f*][1,4]oxazepine derivatives as potent agonists of the human TRPA1 receptor. *Molecules*, **2010**, *15*, 9364-9379.

[113] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G. Co-evolutions of correlations for QSAR of toxicity of organometallic and inorganic substances: An unexpected good prediction based on a model that seems untrustworthy. *Chemom. Intell. Lab. Syst.*, **2011**, *105*, 215-219.

[114] Lee, J.K.; Song, J.S.; Nam, K.D.; Hahn, H.G.; Yoon, C.N. Substituent effects of thiazoline derivatives for fungicidal activities against *Magnaporthe grisea*. *Pest. Biochem. Physiol.*, **2011**, *99*, 125-130.

[115] Lan, P.; Xie, M-Q.; Yao, Y-M.; Chen, W-N.; Chen, W-M. 3D-QSAR studies and molecular docking on [5-(4-amino-1Hbenzoimidazol-2-yl)-furan-2-yl]-phosphonic acid derivatives as fructose-1,6-biphophatase inhibitors. *J. Comput. Aided Mol. Des.*, **2010,** *24*, 993–1008.

[116] Najafi, A.; Ardakani, S.S. 2D autocorrelation modelling of the anti-HIV HEPT analogues using multiple linear regression approaches. *Mol. Simul.*, **2011**, *37*, 72–83.

[117] Sun, J.R.; Lan, P.; Sun, P.H.; Chen, W.M. 3D-QSAR and Docking Studies on Pyrrolopyrimidine Derivatives as LIM-Kinase 2 Inhibitors. *Lett. Drug Des. Discov.,* **2011**, *8,* 229-240.

[118] Mollahasani, A.; Alavi, A.H.; Gandomi, A.H. Empirical modeling of plate load test moduli of soil via gene expression programming. *Comput. Geotech.*, **2011**, *38*, 281–286.

[119] Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Gini, G. Simplified molecular input-line entry system and International Chemical Identifier in the QSAR analysis of styrylquinoline derivatives as HIV-1 integrase inhibitors. *Chem. Biol. Dug Des.*, **2011**, *77*, 343–360.

[120] Toropova, A.P.; Toropov, A.A.; Diaza, R.G.; Benfenati, E.; Gini, G. Analysis of the co-evolutions of correlations as a tool for QSAR-modeling of carcinogenicity: an unexpected good prediction based on a model that seems untrustworthy. *Cent. Eur. J. Chem.*, **2011**, *9*, 165-174.

[121] Zandi-Atashbar, N.; Hemmateenejad, B.; Akhond, M. Determination of amylose in Iranian rice by multivariate calibration of the surface plasmon resonance spectra of silver nanoparticles. *Analyst*, **2011**, *136*, 1760-1766.

[122] Toropov, A.A.; Toropova, A.P.; Lombardo, A.; Roncaglioni, A.; Benfenati, E.; Gini, G. CORAL: Building up the model for bioconcentration factor and defining it's applicability domain. *Eur. J. Med. Chem.*, **2011**, *46*, 1400-1403.

[123] Gandomi, A. H.; Alavi, A. H.; Mirzahosseini M. R.; Nejad, F. M. Nonlinear genetic-based models for prediction of flow number of asphalt mixtures. *J. Mater. Civil Eng.*, **2011**, *23*, 248-263.

[124] Pouretedal H.R.; Keshavarz, M.H. Prediction of toxicity of nitroaromatic compounds through their molecular structures. *J. Iran. Chem. Soc.,* **2011**, *8*, 78-89.

[125] Benfenati, E.; Toropov, A.A.; Toropova, A.P.; Manganaro, A.; Gonella Diaza, R. CORAL software: QSAR for anticancer agents. *Chem. Biol. Drug Des.*, **2011**, *77*, 471–476.

[126] Hao, M.; Li, Y.; Wang, Y.; Zhang, S. Prediction of P2Y12 antagonists using a novel genetic algorithm-support vector machine coupled approach. *Analytica. Chimica. Acta.*, **2011**, *690*, 53–63.

[127] Basu, A.; Sinha, B.N.; Saiko, P.; Graser, G.; Szekeres, T. N-Hydroxy-N'-aminoguanidines as anti-cancer lead molecule: QSAR, synthesis and biological evaluation. *Bioorg. Med. Chem. Lett.*, **2011**, *21*, 3324-3328.

[128] Pissurlenkar, R.R.S.; Khedkar, V.M.; Iyer, R.P.; Coutinho, E.C. Ensemble QSAR: A QSAR method based on conformational ensembles and metric descriptors, *J. Comput. Chem.*, **2011,** *32*, 2204–2218.

[129] Roy, P.P.; Kovarich, S.; Gramatica, P. QSAR Model Reproducibility and Applicability: A Case Study of Rate Constants of Hydroxyl Radical Reaction Models Applied to Polybrominated Diphenyl Ethers and (benzo-) triazoles. *J. Comput. Chem.*, **2011**, *32*, 2386–2396.

[130] Jalali-Heravi, M.; Ebrahimi-Najafabadi, H. Modeling of retention behaviors of most frequent components of essential oils in polar and non-polar stationary phases. *J. Sep. Sci.* **2011**, *34*, 1538–1546.

[131] Wang, S.T.; Ai, Y.; Tang, C.; Song, F.J.; Sun, P.H. 3D-QSAR and docking studies of 4-morpholinopyrrolopyrimidine derivatives as potent mTOR inhibitors. *Lett. Drug Des. Dicov.*, **2011**, *8*, 339-354.

[132] Zhang, H.X.; Li, Y.; Wang, X.; Wang, Y.H. Probing the structural requirements of A-type Aurora kinase inhibitors using 3D-QSAR and molecular docking analysis. *J. Mol. Model*., **2012**, *18*, 1107-1122.

[133] Sharma, B.K.; Singh, P.; Pilania, P.; Shekhawat M.; Prabhakar, Y.S. QSAR of 2-(4-methylsulphonylphenyl)pyrimidine derivatives as cyclooxygenase-2 inhibitors: simple structural fragments as potential modulators of activity. *J Enzym. Inhib. Med. Chem.*, **2012**, *27*, 249-260.

[134] Ojha, P.K.; Roy, K. Exploring QSAR, pharmacophore mapping and docking studies and virtual library generation for cycloguanil derivatives as *Pf*DHFR-TS inhibitors. *Med. Chem.*, **2011**, *7*, 173-199.

[135] Roy, K.; Das, R.N. On some novel extended topochemical atom (ETA) parameters for effective encoding of chemical information and modelling of fundamental physicochemical properties. *SAR QSAR Environ. Res.*, **2011**, *22*, 451-472.

[136] Kar, S.; Roy, K. Development and validation of a robust QSAR model for prediction of carcinogenicity of drugs. *Indian J. Biochem. Biophysics.*, **2011**, *48*, 111-122.

[137] Mitra, I.; Saha, A; Roy K, Chemometric QSAR modeling and *in silico* design of antioxidant NO donor phenols. *Sci. Pharm.* **2011**, *79*, 31-57.

[138] Mitra, I.; Saha, A; Roy K, QSPR of antioxidant phenolic compounds using quantum chemical descriptors. *Mol. Simul.,* **2011**, *37*, 394-413**.**

[139] Ojha, P.K.; Mitra, I.; Das, R.N.; Roy, K. Further exploring r_m^2 metrics for validation of QSPR models. *Chemom. Intell. Lab. Syst.*, **2011**, *107*, 194–205

Received: December 01, 2010 Revised: February 12, 2011 Accepted: February 13, 2011