Medicinal Chemistry

Research Article Open Access

Molecular Modeling Study of the Penetration Kinetic of Diverse Compounds through the Human Skin by Three-Dimensional Quantitative Structure Activity Relationship

Soheila Rezaei[1,2], Hassan Behnejad[1]* and Jahan B Ghasemi[2]

[1]Department of Physical Chemistry, School of Chemistry, University College of Science, University of Tehran, Tehran 14155, Iran
[2]Department of Analytical Chemistry, School of Chemistry, University College of Science, University of Tehran, Tehran 14155, Iran

Abstract

The control of permeation is essential for the topical application of lotions, creams, and ointments, and the toxicological and risk assessment of materials from environmental and occupational hazards. This study developed a three-dimensional quantitative structure-activity relationship (3DQSAR) model to predict permeation of a variety of 210 compounds through human skin. Molecular descriptors were computed using a GRid Independent Descriptors (GRIND) approach. After variable selection via the genetic algorithm method, the 118 selected descriptors were correlated with skin permeability constants by PLS regression and support vector machine (SVM). Partial least squares regression (PLSR) and support vector regression (SVR) are two popular Chemometrics models that are being subjected to a comparative study in the presented work. Kennard-Stone algorithm was employed to split data set to a training set of 150 molecules and a test set of 60 molecules. Genetic algorithm (GA), as an influential linear tool, was used to certain the best and interpretative subset of variables for the predictive model structure. The best results were obtained by PLS regression with the correlation coefficient of R^2=0.77 and SVM regression with the correlation R^2=0.79. This strategy led to a final 3DQSAR model that presented Q^2=0.61 and R^2 pred=0.73. The obtained results revealed that the hydrogen bonding donor and hydrogen bonding acceptor of investigated compounds dramatically influences their ability to penetrate through human skin. Furthermore, it was found that permeability was enhanced by increasing hydrophobicity and lowered with increasing molecular weight.

Keywords: Skin permeability; 3D-QSAR; GRIND; Molecular weight; Log k_p; PLS

Abbreviations: k_p: Epidermal Permeability coefficient; J_{ss}: Steady-state flux; R^2: Determination coefficient; C_d: Chemical concentration in dose formulation; D: Diffusion coefficient in the skin; GRIND: Grid Independent Descriptors; K: Skin vehicle partition coefficient; L: Thickness of skin; LV: Latent variable; MIF: Molecular interaction field; MW: Molecular weight; PLS: Partial least squares; QSAR: Quantitative structure-activity relationship; RMSEC: Root mean square error of calibration; RMSECV: Root mean square error of cross-validation; RMSEP: Root mean square error of predication; SVR: Support vector regression.

Introduction

3D-QSAR methods are now standard tools in medicinal chemistry projects. The Grid-Independent Descriptors (GRIND) was published in the year 2000 and the first version of the software used to generate GRIND-based 3D-QSAR models (ALMOND) was available in the same years [1,2]. Briefly, the calculation of GRIND descriptors involves three steps: computing a set of molecular interaction fields (MIFs) for molecules in the data set, filtering the MIFs to extract the most relevant regions, and encoding the filtered MIFs (also called final nodes where a node represents a favourable probe target molecule interaction region) into GRIND variables. This procedure works on the final nodes and computes the product of the interaction energy for each pair of nodes. The products are then ranked according to the distance between nodes. Distances are grouped in a discrete number of categories and in each category, only the product with the highest value is stored and represent one GRIND variable. GRIND variables are then grouped into blocks (=correlogram) representing distances between pairs of nodes. Finally, the GRIND variables constitute a matrix of descriptors that can be analysed using multivariate techniques [1].

Three main reasons underlie the appeal of these descriptors: GRIND is alignment-independent, they are chemically interpretable and quick and easy to compute. GRIND-based 3D-QSAR models have thus been successfully used to describe some biological topics [3-5].

Quantitative structure-activity/property relationship (QSAR/QSPR) approaches, as progressive tools in modeling and prediction of many physiochemical properties offer a fast measure of predictability in the absence of extensive experimental or computed data on compounds properties. 3D-QSAR, which refers to use of force field calculations to compute spatial properties of a three-dimensional structure (3D) of compounds, provide useful information of the forces and interactions between two molecules [6,7]. The GRIND, alignment independent, interpretable and efficient to compute descriptors derived from GRID molecular interaction fields, was proved relevant in diverse structure-activity relationship studies.

The skin is the human body's largest organ and protects the body from the xenobiotic influx. Local and systemic drugs can be distributed throughout the skin. Prediction of skin permeability is also crucial in toxicological assessment after topical exposure. Predicting human skin permeability of chemical compounds accurately and efficiently is useful for developing dermatological medicines and cosmetics. The skin permeability of a solute depends on several parameters, namely, chemical concentration in dose formulation (C_d), skin-vehicle partition

***Corresponding author:** Hassan Behnejad, Department of Physical Chemistry, School of Chemistry, University College of Science, University of Tehran, Tehran 14155, Iran, Tel: +98 21 61113639; E-mail: behnejad@khayam.ut.ac.ir

Received June 05, 2019; **Accepted** June 17, 2019; **Published** June 24, 2019

Citation: Rezaei S, Behnejad H, Ghasemi JB (2019) Molecular Modeling Study of the Penetration Kinetic of Diverse Compounds through the Human Skin by Three-Dimensional Quantitative Structure Activity Relationship. Med Chem (Los Angeles) 9: 065-073.

coefficient (K), diffusion coefficient in the skin (D), and thickness of the skin (L). The permeability coefficient (k_p) quantifies the percutaneous absorption of chemicals through the skin defined as follows [8]:

$$k_p = \frac{K.D}{L} = \frac{J_{ss}}{C_d} \quad (1)$$

Where, J_{ss} is the steady state flux of the solute.

Early QSARs to predict skin permeation of chemicals revealed that hydrophobicity was correlated linearly with increasing permeability [9,10]. Using a larger dataset, Flynn presented a QSAR approach for 95 chemicals, reporting algorithms for compounds with low and high molecular weights. Also, they demonstrated that skin permeability (k_p) was a function of partitioning between aqueous and non-aqueous layers, as described by the octanol-water partition coefficient. Subsequently, many QSARs for skin permeation have been published utilizing the Flynn dataset, where the data have either been used as a whole or subsets taken into account for different types of compounds [11]. Potts and Guy were among the first to demonstrate the use of the logarithm of the octanol-water partition coefficient ($\log k_{ow}$) in combination with molecular size descriptors, i.e., molecular weight (MW) or molecular volume (MV) to model skin permeability [12,13]. They obtained the following relationship as follow:

$$\log k_p = 0.711 \log k_{ow} - 0.0061\ MW - 6.3 \quad (2)$$

Other QSAR analyses have confirmed hydrophobicity and molecular size to be the influential descriptors of skin permeability for chemicals. It has also been demonstrated that the hydrogen bonding character of a compound might also influence greatly its ability to penetrate through the human skin [14-21]. Hydrogen bonding has been parameterized using a number of approaches including the number of hydrogen bonds that may be formed by a compound (H_b) hydrogen bond donor (H_d) and acceptor (H_a) ability of a compound, and the HYBOT plus series of descriptors introduced by Raevsky et al., as well as many others [20,22,23].

To achieve this goal, a dataset of 210 compounds, which had not previously been utilized for QSAR development, was analyzed. Furthermore, to explore the fundamental physicochemical meaning of skin permeability, a whole variety of physicochemical properties, hypothesized to be of significance, were calculated. These included calculated descriptors unique to this study to investigate the role of molecular size and hydrophobicity in skin permeability. Results indicate that GRIND-based 3D-QSAR approach can reliably predict continuation of our data in different areas in chemistry and related fields using the QSAR approach, is to develop, for the first time to our knowledge, a valid and predictive 3D-QSAR model, able to correlate and predict skin permeability with the applicability of GRID independent descriptors (GRIND) [24,25].

Materials and Methods

Data set

A new large dataset of 210 structurally diverse compounds with human skin permeability coefficients presented in the supplementary material. The dataset was compiled which are not obtained by *in vitro* diffusion study [26]. Therefore, this database is highly appropriate for evaluating the relationships between $\log k_p$ and human skin permeability. It is hypothetically applicable to related research such as vehicle effects on skin permeability by using it combined with a dataset of skin permeability for solutions other than water. They were working on compiling a skin permeability dataset composed of a wide variety of permeants and solvents to develop the prediction models of vehicle effects on skin permeability in order to optimize topical formulations.

Our database has experimental skin permeability coefficients determined from the excised human skin. This database does not contain calculated permeability, data from other animals and data with chemical or physical penetration enhancement. Therefore, from this database, we can build models that directly detect the influence of molecular structures on human skin permeability. Our high-performance GRIND prediction models will be reliable and useful tools for developing dermatological ingredients.

Compounds in this extensive database belong to various chemical classes, including alcohols, amines, amides, aromatics, carbonyls, carboxylic acids, esters, ethers, urea, halides, nitriles, and nitro compounds. Many of the compounds are active ingredients of pharmaceutical products, such as anti-inflammatory, anti-cancer, anti-HIV, local anesthetic, stimulants, and sleep-inducing drugs.

Subset selection

A moderated Kennard-Stone algorithm, where the response vector has been replicated k (number of descriptors) times to enhance the influence of the response on the splitting results, was employed to split the data set into training and a prediction set [27]. The training set of 150 molecules was used to adjust the parameters of the QSAR model, and the rest were used to evaluate models prediction ability as a test set (60 compounds).

Molecular optimization and descriptor calculation

The structure of molecules was drawn in ChemBioOffice 11.0 and the mol files format exported in HyperChem 8.0.5. Two stages of molecular orbital (MO) calculation including MM+ force field and semi-empirical method AM1, gradient norm criterion 0.01 kcal/Å, were applied in the geometry optimization for all structures.

Grid-INdependent descriptors are a new class of molecular descriptors developed by Pastor et al. [5]. Pentacle software 1.05 (Molecular Discovery Ltd, Oxford, UK) has been proposed as a tool to extract descriptors [28]. Grid molecular interaction fields (MIFs) of nodes are computed by four GRid probes, and a pair of nodes (GRid MIF minima) is used as descriptors (variables) [29]. Only those pairs of nodes (for the same or different probe types) with the highest product of interaction energy (IE), at the given distance range, were used for the PLS analysis. For the derivation of MIFs, four most recommended probes were used. To represent steric and hydrophobic interactions, hydrogen bond acceptor, and hydrogen bond donor groups, we used DRY (hydrophobic probe), O (carbonyl oxygen), and N1 (amide nitrogen), respectively. These probes stand for strong non-covalent interactions between molecules and receptor. Moreover, to regard molecular shape effects in the receptor-ligand interaction process, and as complementary to point interaction based information, a supplementary probe, called TIP (shape probe), was applied that extracts each ligand's isosurface at 1 kcal/mol from the field of a normal GRid calculation. AMANDA algorithm as implemented in the software was applied for the filtering. This algorithm is the regions with the most relevant MIF that uses the intensity of the field at a node and the mutual node-node distances between the chosen nodes [30]. At each point the interaction energy (E_{xyz}) was calculated as a sum of Lennard-Jones energy (E_{lj}), hydrogen bond (E_{hb}) and electrostatic (E_{el}) interactions.

$$E_{xyz} = \sum E_{lj} + \sum E_{el} + \sum E_{hb} \quad (3)$$

Maximum auto and cross-correlation (MACC-2) algorithm were applied for the encoding. The grid spacing was set to 0.5 Å and the

smoothing windows to 0.8 Å. The MACC-2 (Maximum Auto and Cross Correlogram) analysis output is usually represented directly in correlograms where each point represents the product of two particular nodes within the distance box separating the nodes of a certain compound. The values obtained from this analysis were represented directly in correlogram plots, where the product of node-node energies is reported versus the distance separating the nodes. Highest energy product can be defined for the same probe (obtaining four auto correlograms: DRY-DRY, O-O, N1-N1 and TIP-TIP) and for pairs of different probes (obtaining six cross correlograms: DRY-O, DRY-N1, DRY-TIP, O-N1, O-TIP, and N1-TIP).

Each block (or correlogram) of variables corresponds to a type of interaction between a couple of nodes: DRY (which represent hydrophobic interactions), O (SP^2 carbonyl oxygen, representing H-bond acceptor), N1 (neutral flat NH, like in amide, H-bond donor) and the TIP probe (molecular shape descriptor).

Variable selection and modeling

Several methods can be applied to decrease the original pool of descriptors to an appropriate size and to select the most informative variables. Reducing the pool of descriptors eliminates those descriptors that contribute either no information or whose information content is redundant with other descriptors present in the pool [31].

In present work, we applied the genetic algorithm (GA) variable selection method to GRIND. The GA method with a well-chosen objective function outclasses the traditional approaches and it is a superior option to achieve a representative subset of variables from a multi-experiment dataset [32]. GA is a method for moving from one population of chromosomes to a new population by using a kind of natural selection together with the genetics, inspired operators of crossover, mutation, and inversion [33]. The selected descriptors were employed to generate the models with the PLS approach. The flexibility of the PLS approach, its graphical orientation, its essential ability to handle incomplete noisy data with many variables and observations make PLS as a simple but powerful approach for the analysis of data in complicated problems [34,35]. For a dataset ($X_{210 \times 740}$) containing 210 compounds, 740 descriptors have been obtained, using four GRID probes. After removing the descriptors containing only zero or constant values for all solute and the remaining 476 ($X_{210 \times 476}$), then the genetic algorithm (GA) was used to extract the more informative variables and generate the more predictive model with 118 descriptors ($X_{210 \times 118}$).

The GA was carried out during 200 generations with 30 chromosomes in 1,000 runs with a probability of mutation 1% and the probability of cross-over 50% using the PLS-genetic algorithm toolbox. All descriptors and target values were normalized between 0 and 1 prior to network training. All calculations were performed in the MATLAB (version 7.6.0, Math Works, Inc.).

Results and Discussion

It is widely believed that 3D descriptors should provide better descriptions of interactions between two compounds. However, most 3D methods suffer from two constraints: use the correct conformation of a molecule, which may not even be the lowest energy conformation to compare structurally different compounds, and proper alignment of the compounds, a step that is time-consuming and may introduce user bias [36]. GRIND procedure was developed with the aim to overcome the alignment problem and was therefore selected for this study. This method is GRID-based MIFs that calculates the interaction energies between the molecule and chemical probes. When MIFs are computed for a molecule, the region showing favourable energies of interaction represent positions where two molecules would interact favourably with each other.

In a few words, the GRIND methodology involves three steps: computing a set of molecular interaction fields (MIFs) for molecules in the data set, filtering the MIFs to extract the most relevant regions, and encoding the filtered MIFs into molecular descriptors named GRIND. These descriptors represent the most important GRID interactions as a function of the distance instead of the position of each grid point [5,37-39].

Model construction

The chemometric analysis was carried out using the statistical tools included in Pentacle software. The GRIND descriptors were related to stability constants through partial least square (PLS) analysis. We selected the model with the best statistical performance. The algorithm of the used PLS was SIMPLS. It is precisely the same as PLS when there is only one response and invariably gives very similar results, but it can be dramatically more efficient to compute when there are many factors [40].

Table 1 displays the statistical results of different models. We selected the model with the best statistical performance. The best model was PLS model with six latent variables in Table 2 that selected on the basis of the highest squared correlation coefficient (R^2), Equation 3, cross-validated squared correlation coefficient (Q^2), Equation 5, and simplicity and interpretability of the model. The most popular measures of how well a model fits the data are probably the mentioned above parameters. The R^2, Q^2 and RMSEC (Root mean square error of Calibration) (Equation 6) for training set were (0.77, 0.61, 0.49) and R^2, and RMSEP for test set were (0.73, 0.61), respectively.

$$R^2=1-\frac{\sum_i (y_i^{cal}-y_i^{obs})^2}{\sum_i (y_i^{obs}-y^{mean})^2} \quad (4)$$

$$Q^2=1-\frac{\sum_i (y_i^{pred}-y_i^{obs})^2}{\sum_i (y_i^{obs}-y^{mean})^2} \quad (5)$$

$$RMSEC=\sqrt{\frac{\sum_i (y_i^{cal}-y_i^{obs})^2}{N}} \quad (6)$$

RMSEP (Root mean square error of prediction) is calculated exactly as in Equation 6 except that the estimates y_{cal} are based on a previously developed model, not one in which the samples to be 'predicted' are included in the model building.

Model validation

The final generated QSAR model was validated on the basis of an

Method	R^2_{cal}	Q^2	R^2_{pred}	RMSEC	RMSECV	RMSEP
PLS	0.77	0.61	0.73	0.49	0.66	0.61
SVM	0.79	0.50	0.79	0.01	0.73	0.50

Table 1: Statistical results of different methods for correlation log k_p to grind descriptors.

LV	R^2_{cal}	Q^2	R^2_{pred}	RMSEC	RMSECV	RMSEP
1	0.35	0.31	0.57	0.84	0.86	0.84
2	0.51	0.42	0.67	0.73	0.79	0.73
3	0.69	0.54	0.67	0.58	0.71	0.74
4	0.72	0.58	0.71	0.54	0.67	0.64
5	0.75	0.61	0.75	0.51	0.65	0.6
6	0.77	0.61	0.73	0.49	0.66	0.61

Table 2: Statistical results of PLS models for the first 6 LVs.

external test set consisting of 60 compounds with known experimental log k_p values but was not used in the training set. In addition to the external validation, the model was further optimized by an internal validation. Leave-one-out cross-validation is a particular case of k-fold cross-validation where k equals the number of instances in the data. In other words, in each repetition nearly all the data except for a single observation are used for training and the model is tested on that single observation [41,42].

In PLS modeling, we assume that the investigated system or process actually is influenced by just a few underlying variables, latent variables (LV's). The number of LV's is usually not known. This number is generally chosen by cross-validation considering the proportion of variations explained by each latent variable [34]. Figure 1 reveals that six LVs were found to be significant for internal validation using cross-validation (Q^2=0.61). The plot of experimental and predicted log k_p values for PLS model is shown in Figure 2. In this figure the red trace is 1:1 line (the best fit mode) and the purple line is the fit (regression) line which shows the R^2 train (calibration set) value.

Support vector machine (SVM) is a classification algorithm, which has been widely used in machine learning and in silico prediction because of its remarkable versatility. The theory of SVM is detailed in several excellent sources [43,44]. The key point of SVM is kernel transformation that is a projection of the descriptor matrix from the input space into a high dimensional feature space. In this study, we used the radial basis function (because such kernels are standard and well used) and the option called epsilon-type regression (eps-regression) and gamma. We used default settings for all tunable parameters in the SVM function. The primary values for ε was 0.1 and for γ 0.0001. The plot of experimental and predicted log k_p values for the SVM model is shown in Figure 3. In comparing the two models, significant differences between statistical parameters by linear PLS and nonlinear SVM, based on the same input variables, were not observed. The predicted values were listed in Table 1 for both PLS and SVM models. All calculations were performed in the MATLAB and the PLS Toolbox 5.8.2 (Eigenvector Research, Inc., Manson, Washington, USA) media.

Applicability Domain (AD)

A predicted value without an idea of the reliability of the value is not useful when you have a new compound with no experimental data. Therefore, in order to use a QSAR model for evaluating new compounds, its domain of application needs to be defined. Only those predictions that lie within this domain may be considered as reliable.

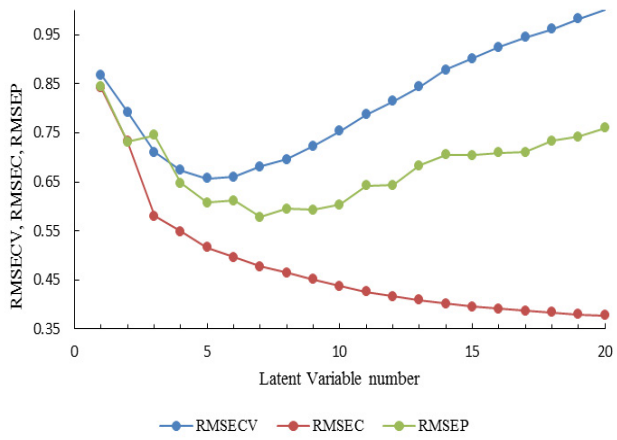

Figure 1: SIMPLS variance captured versus latent variable number.

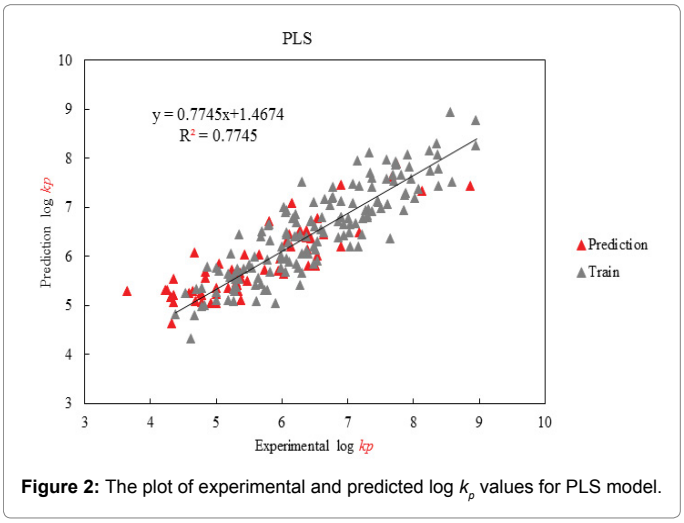

Figure 2: The plot of experimental and predicted log k_p values for PLS model.

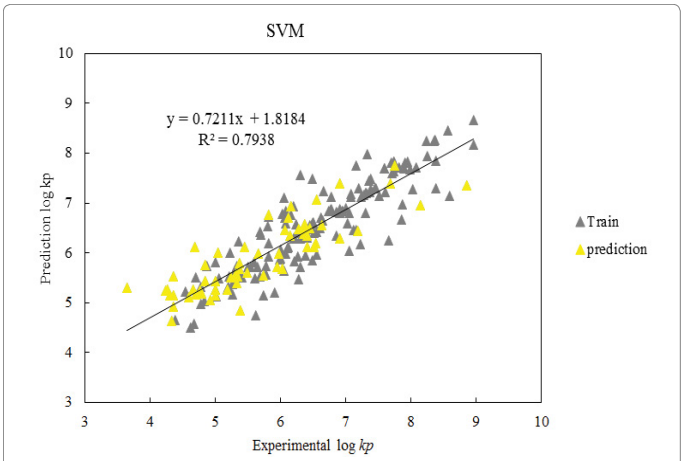

Figure 3: The plot of experimental and predicted log k_p values for SVM model.

The extent of extrapolation is a simple measure to define the applicability domain. It is based on the calculation of the leverage hi for each compound, where the QSAR model is used to predict its activity. The leverage of a compound is a measure related to the statistical error of prediction of that compound. Equation 7 shows the leverage for objects in the PLS model as follow:

$$h_{ii} = t_i^T \times \left(T^T \times T\right)^{-1} \times t_i \quad (7)$$

where h_{ii} leverage of the i^{th} sample is the i^{th} diagonal element of the scores matrix T which is truncated according to number of significant factors. The studentized residual can be calculated by Equation 8;

$$r_i = \frac{e_i}{RMSEC\sqrt{1-h_{ii}}} \quad i=1,2...n \quad (8)$$

Where e_i the residual of the i^{th} object and the RMSEC is presented in Equation 6.

A leverage value h_i>2 (k+1)/n is considered large where k is the number of model parameters plus one and n is number of training set molecules. This criterion means that the predicted response is the result of a substantial extrapolation of the model and may not be reliable [45-48].

In the present case, the training descriptor matrix X was of order 150 × 118 and thus the threshold leverage value was 0.09. The results indicate that out of the 60 test compounds (not present in the training

set). Thus, 90% of the test compounds are within the applicability domain indicating that their predicted activity values are reliable. The analysis of the applicability domain of the PLS model based on GRIND descriptors displayed in Figure 4. As obvious, there is 10 chemical outlier or structure influential compound in the training set and 12 chemical outlier in the testing set. compounds such as Sucrose (M200), Raffinose (M195), Hydrocortisone succinate (M122) and Ouabain (M178), in training set due to their high molecular weight and having strong alcoholic, etheric and ester groups, caused the lack penetration, and then these compounds, according to the application domain, are outlier and more than h* which helps us eliminate these compounds to build of the better model. The Ouabain (M178) is the most inactive molecule in the penetration, and the Williams plot is well illustrated and depicted it as a distorted compound.

It is also important to note that the validation chemicals, which were not used for model development, are predicted with similar accuracy as the training chemicals. It was found that majority of these compounds have moderate and high permeation level. It is believed that keeping these samples in the training set could significantly worsen the model statistic and they should be deleted from the training set [49]. However, it has been discussed that such compounds may hold unique information, which makes the model more precise and that their removal should be carefully considered [50]. Figure 4 reveals the presence of just one chemical outlier M204 (thymol) with studentized residual more than standard deviation units (>3 s) that were considered as y outlier but incorrect within the model domain.

Descriptors interpretation

Analysis of the PLS coefficients profile of the GRIND model allows identifying those descriptors which exhibit the largest contribution to the model. Figure 5 shows the PLS coefficient plot indicated the most important pairs of nodes that contribute negatively or positively to the permeation in pentacle with 740 descriptors. A first inspection of the PLS coefficients plot enabled us to select some X variables with the highest impact on the Y variable.

The PLS coefficient histogram showing the contribution of every single variable to the model versus the value of Y (118 descriptors) is shown in Figure 6. Positive values of the coefficients represent a direct correlation to the Y, and the negative ones show an inverse correlation to it. As can be seen, the most important variables that have a positive effect on the permeation activity are O-TIP: 606,607 N1-N1: 170, O-O: 104, and DRY-DRY: 20, 21. In contrast, the analysis of all the distances at higher PLS coefficients revealed that the variables TIP-TIP: 231, O-TIP: 634 and DRY-TIP: 462 correlate negatively with the permeation activity. In the other word, these probes are participating most in explaining the variance in the permeation activity values (Figure 6).

The largest peaks were related to the TIP probe (correlograms O-TIP, TIP-TIP, and DRY-TIP), which represent shape and size of the molecules suggesting that size and shape of the molecules likewise presence and orientation of hydrogen bond groups were determinant for the penetration of the skin. The next effective peaks were related to the O probe correlograms (O-O, O-TIP, and O-N1), which represent bond acceptor groups of the molecules whereas N1 and DRY probes showed significant peaks also.

In the primary analysis, in addition to the shape and size of the molecular orientation of the hydrogen bond, it is essential to penetrate these compounds. To get deeper results for the QSAR model, the variables with the highest impact on skin permeation are shown in more detail in Table 3. The chemical interpretation of the model was

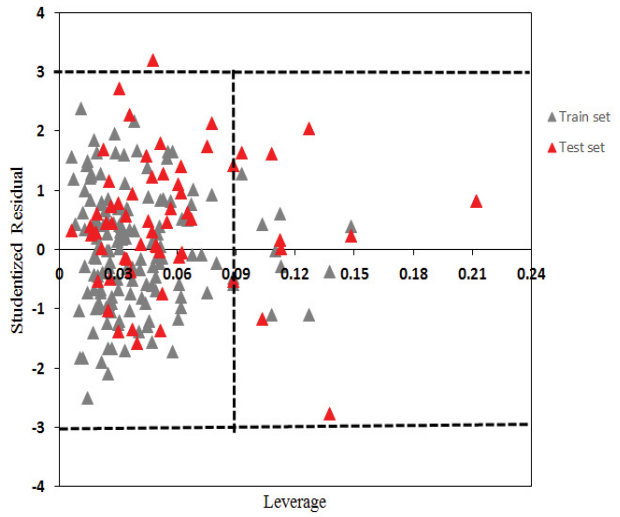

Figure 4: Plot of studentized residuals versus leverages. Dotted lines represent ± 3 studentized residuals and dash line represents warning leverage (h∗ ≈ 0.09).

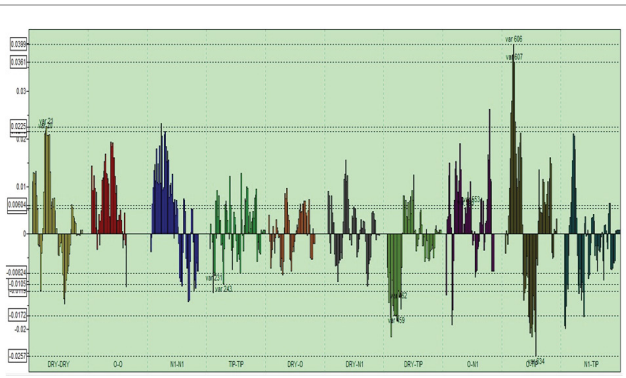

Figure 5: PLS Coefficient plot (All Descriptors).

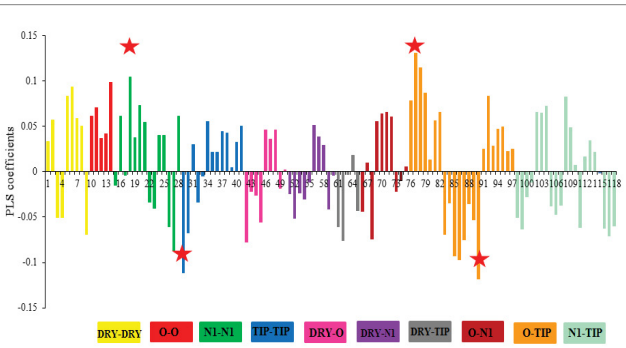

Figure 6: PLS Coefficient plot (118 Descriptors).

assessed by selecting the twenty most relevant descriptors: three O-N1 probes, two DRY-DRY probes, one O-O prob, two N1-N1 probes, two TIP-TIP probes, one DRY-O probe, two DRY-TIP probes, five O-TIP probes, and two N1-TIP probes.

Variable number 77 in extracted PLS coefficient in Figure 6 that refers to variable O-TIP 606, with distance: 5.60-6 Å is the largest impact on skin permeability with a direct relationship. Great size molecules like: M178, M207, M154, and M209 (Ouabain, Triamcinolone

No. of variable	Prob	Distance (A°)	Coefficient sign
5	DRY-DRY	8-8.4	+
6	DRY-DRY	8.4-8.80	+
14	O-O	12-12.40	+
18	N1-N1	8.80-9.20	+
27	N1-N1	20.80-21.20	-
29	TIP-TIP	3.60-4	-
30	TIP-TIP	8.40-8.80	-
42	DRY-O	2.80-3.20	-
61	DRY-TIP	6-6.40	-
62	DRY-TIP	7.20-7.60	-
68	O-N1	4.80-5.20	-
70	O-N1	11.20-11.60	+
71	O-N1	14-14.40	+
77	O-TIP	5.60-6	+
78	O-TIP	6-6.40	+
85	O-TIP	14.40-14.80	-
86	O-TIP	15.20-15.60	-
90	O-TIP	16.80-17.20	-
103	N1-TIP	6.80-7.20	+
104	N1-TIP	7.20-7.60	+

Table 3: The most relevant GRIND descriptors.

acetonide, Morphine and Triglycol nicotinate) with molecular weight respectively: (584.73, 434.55, 285.37, 255.3) had the largest value of variable 77, but for a medium size molecule like M60 (butyl nicotinate) with 179.24 molecular weight has a low value of that variable. For example Triamcinolone acetonide has not any significant permeation enhancement. This variable demonstrated that the hydrogen bonding properties of a compound and shape of the molecules might also influence greatly its ability to penetrate through human skin, this prob also shown that solute hydrogen bonding acceptor may impede diffusion in the stratum corneum. Hydrogen bonding acceptor has also been suggested as a determinant of epidermal penetration, with the numbers of hydrogen bonds. Furthermore, Figure 7 shows a graphical display of variable 77 for the most inactive molecule (M178: Ouabain) and active molecule such (M168: n-octanol). Due to possessing hydrogen bond acceptor group and shape of the molecule, (M33: acetic acid) with (log k_p=6.08) strongly exhibits positive interaction with probe O-TIP. This descriptor indicates that molecular size and the hydrogen bonding capability of a molecule affect its ability to permeate of the skin. Figure 8 shows a graphical display of variable 77 for (M33: acetic acid). We find out from Table 3 that N1-TIP and O-TIP cross-correlogram are several variables of high intensity that have a significant impact on the model.

As N1 probe represents hydrogen bonding (HB) acceptor interaction, it is important in compounds with hydrogen bond donor group. The graphical display of variable 18 that refers to variable 170 for selected compound (M64: Caffeine) was shown in Figure 9. Variable 18, N1-N1 at distance of 8.80-9.20 Å, which has a positive correlation with permeation constant indicated that the hydrogen bonding capability of a molecule affects its ability to permeate the skin. The highest value of energy interaction product for the GRIND variable 18 in compound 64 which is clearly evident in correlogram means that the N1 probe interacts more strongly with the nitrogen atom in Caffeine than the groups described above.

Variable number 90 refer to O-TIP variable 634 that indicated a significant distance of 17.2-17.6 Å between O and TIP nodes, which has a negative correlation with inactive compound, As expected this variable is not expressed for the most active compound. Figure 10

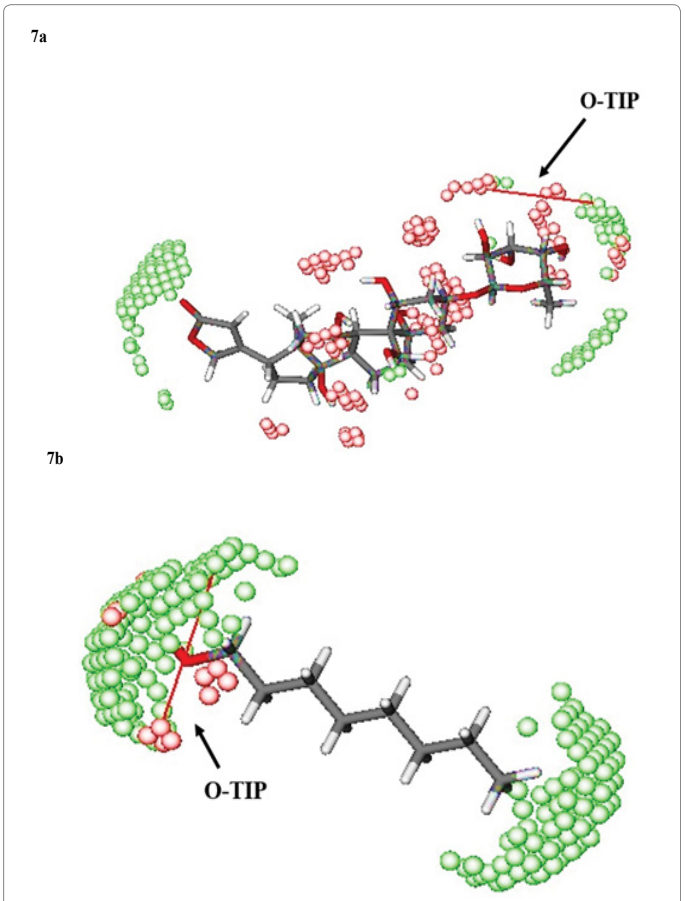

Figure 7: Graphical display of GRIND variable 77 (O-TIP: red and green hot spot) for the most inactive (M178) and the active (M168).

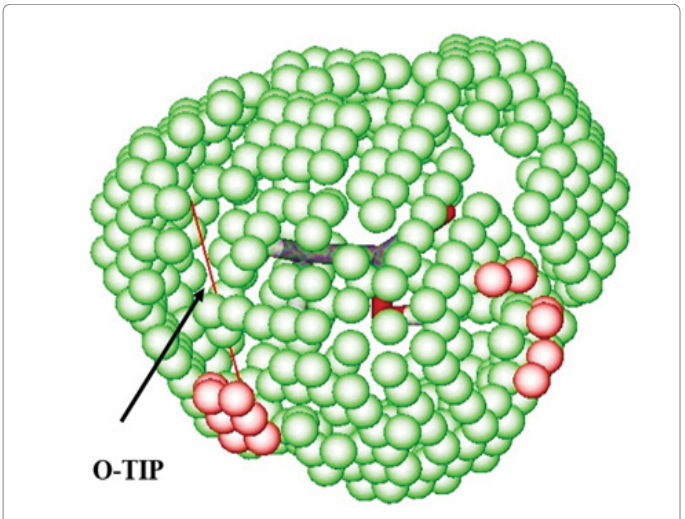

Figure 8: Graphical display of GRIND variable 95 of the (O-TIP) cross-correlogram of a selected (M33) compound.

shows a graphical display of variable 90 for the inactive compound (M178: Ouabain) and active compound (M168: n-octanol).

Number 6 related to GRIND variable 21 which is a DRY-DRY type. The 3D-QSAR model using GRIND descriptors further refines this general property and identified two hydrophobic regions (DRY-

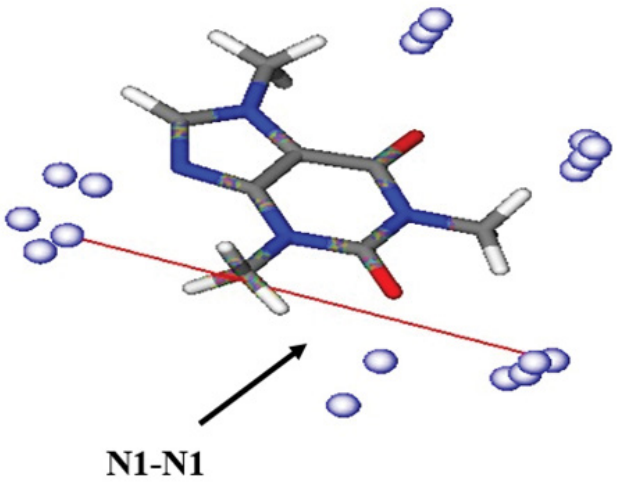

Figure 9: Graphical display of variable 18 (N1-N1: blue hot spot) for selected compound (M64).

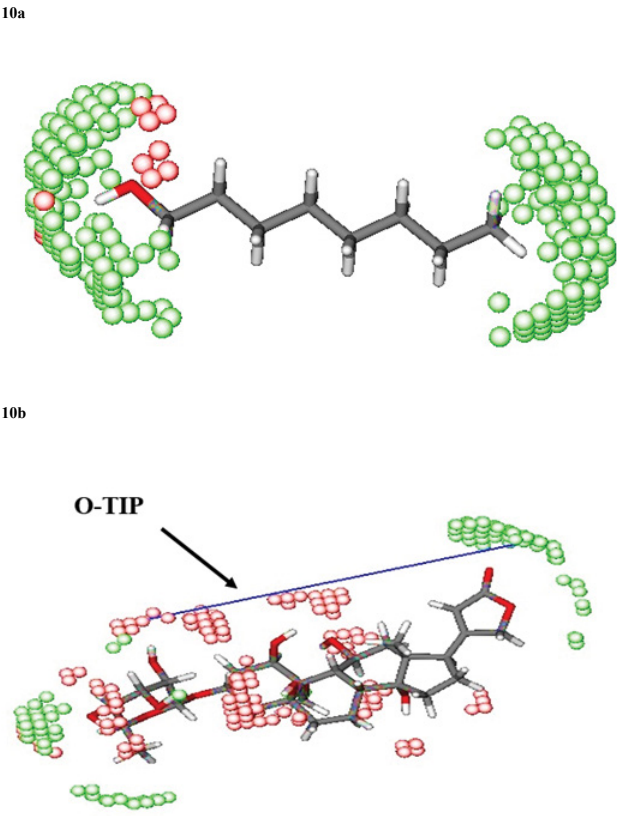

Figure 10: Graphical display of GRIND variable 90 (O-TIP) for the most inactive compound (M178) and active compound (M168).

DRY) separated by a certain distance range in all active compounds. As expected it has a positive effect because as the guest molecules become more hydrophobic, the interaction between them and the hydrophobic cavity becomes stronger. This variable indicated a significant distance of (8.4-8.80 Å) between DRY nodes which have a positive correlation with permeation. This variable was expected to be present in the most active compound 205 and has direct interaction between the aromatic group and the aliphatic group. This variable is present in a selected compound (M61) with a coefficient of permeability of 4.61 as an active compound. As expected this variable has any correlation for a selected inactive compound (M26: 4-hydroxybenzyl alcohol) with a coefficient of permeability of 6.26. According to PLS coefficient plot, most of the variable in the DRY-DRY block has a positive impact on permeation (Figure 6). At a deeper insight, the presence of hydrophobicity groups such as aromatic ring system and long aliphatic chains increases the permeability of the compounds. It is well known that DRY probe has a high affinity to different types of π systems, aromatic moieties or vinyl type. The graphical display of this variable for the selected compound, most active compound and selected inactive compound was shown in Figure 11.

Molecule 61(butyl p- amino benzoate) with log k_p 4.61 and molecular weight 179.24 as active molecule with variables 5, 6 (DRY-DRY), 78(O-TIP) and 103(N1-TIP) have direct effect and positive contribution to permeability. These represent the aromatic ring of the benzoate ring system and hydrogen bond donor of Amin with other

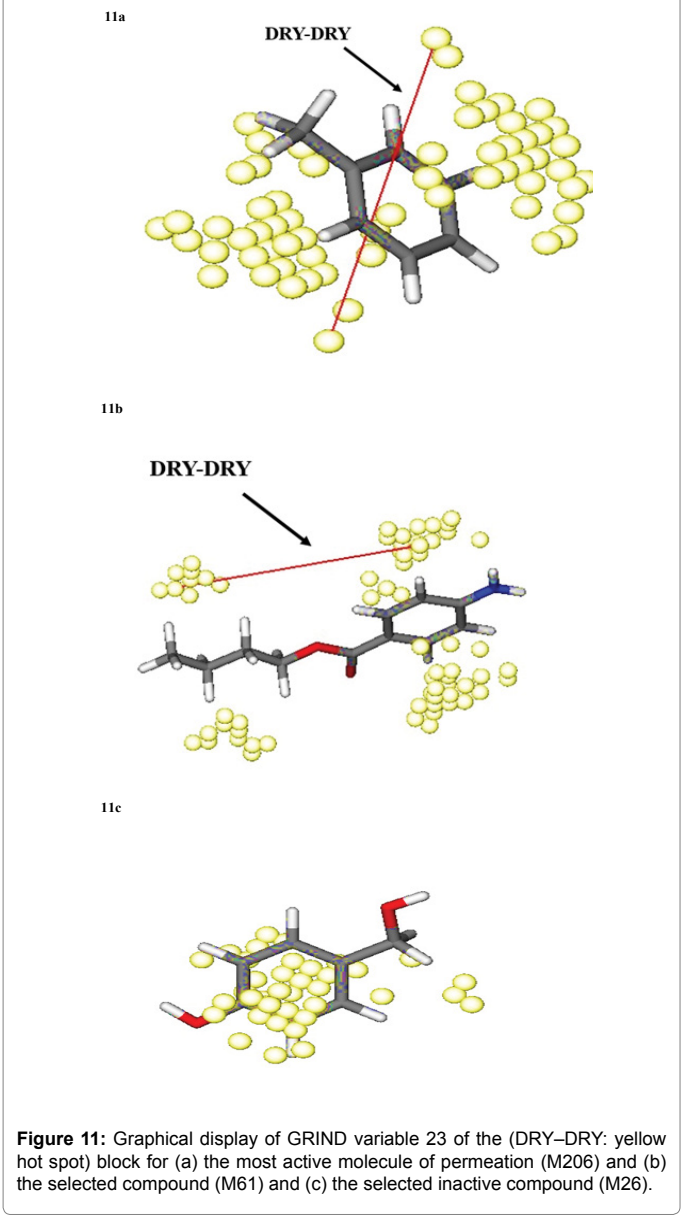

Figure 11: Graphical display of GRIND variable 23 of the (DRY–DRY: yellow hot spot) block for (a) the most active molecule of permeation (M206) and (b) the selected compound (M61) and (c) the selected inactive compound (M26).

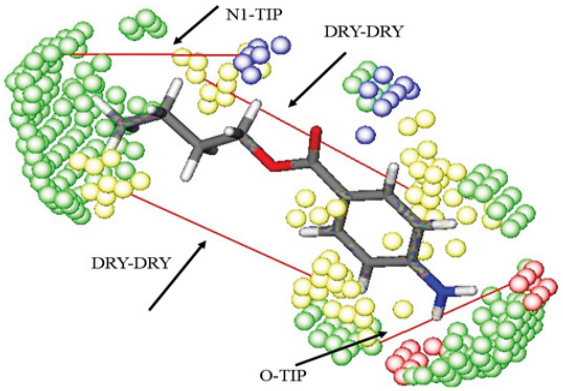

Figure 12: Graphical display of GRIND variables 5, 6 (DRY-DRY), 78 (O-TIP), 103 (N1-TIP) for selected compound (M61).

groups. In the most active compound are separated by a distance of 8.4-8.80 Å and 6.80-7.20 Å, which is considered optimal according to the GRIND model (Figure 12).

According to Figure 6, all O-O interactions demonstrated positively related to permeation. GRIND variable 14 which have the strongest positive impact on permeability are within the block of O-O node pairs. Also this probe demonstrated the presence of hydrogen bonding groups has a direct effect on skin permeability.

Conclusion

The work introduced in this paper aimed to compare two different methods for multivariate calibration, PLS and nonlinear SVM highlighting the underlying algorithm for each and making a modest comparison between them to indicate their merits and demerits. The results of both models performed an acceptable prediction of the model but Values of RMSEP of independent test set (0.61) reveal that linear PLS is better than SVM in prediction ability of the future samples and shows higher generalization ability. GRIND-based 3D QSAR models can give different kinds of information: simple and fast, reliable prediction of activity of compounds belonging to the data set and chemical interpretation of the obtained results. So, in the present work, we investigated the reliability of Grind methodology in predicting human skin permeability data of 210 different compounds acquired from aqueous donors in various literature reports. The hydrophobicity (associated with the DRY probe), shape effects (associated with TIP probe), and hydrogen bond acceptor-donor interactions (associated with N1 probe) are the main factors that determine skin permeability coefficient, within the studied set. GRIND variables, including O-TIP (5.6-6.00 Å) and N1-N1 (8.80-9.20 Å), had greater importance on permeation constant. We illustrated that the hydrogen bonding properties of the investigated compounds greatly influenced its ability to penetrate through human skin. We also found that some descriptors such as DRY and TIP showed the significant peaks which, showed hydrophobicity and molecular size. Furthermore, it was found that permeability was enhanced by increasing hydrophobicity and decreased with increasing molecular weight.

References

1. Cruciani G (2006) Molecular interaction fields: applications in drug discovery and ADME prediction. Wiley-VCH Verlag GmbH & Co. KGaA, Germany.
2. Kastenholz MA, Pastor M, Cruciani G, Haaksma EE, Fox T (2000) GRID/CPCA: a new computational tool to design selective ligands. J Med Chem 43: 3033-3044.
3. Cianchetta G, Li Y, Kang J, Rampe D, Fravolini A, et al. (2005) Predictive models for hERG potassium channel blockers. Bioorg Med Chem Lett 15: 3637-3642.
4. Fontaine F, Pastor M, Zamora I, Sanz F (2005) Anchor grind: filling the gap between standard 3d qsar and the grid-independent descriptors. J Med Chem 48: 2687-2694.
5. Pastor M, Cruciani G, McLay I, Pickett S, Clementi S (2000) GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. J Med Chem 43: 3233-3243.
6. Cruciani G, Carosati E, Clementi S (2003) Three-dimensional quantitative structure-property relationships. In: Wermuth CG (ed.). The practice of medicinal chemistry, Academic Press, New York, USA, pp 405-416.
7. Langer T, Bryant S (2008) 3D quantitative structure-property relationships. In: Wermuth CG (ed.). The Practice of Medicinal Chemistry. 3rd edn. Academic Press, New York, USA, pp. 587-604.
8. Blank IH, McAuliffe DJ (1985) Penetration of benzene through human skin. J Investig Dermatol 85: 522-526.
9. Roberts MS, Anderson RA, Swarbrick J (1977) Permeability of human epidermis to phenolic compounds. J Pharm Pharmacol 29: 677-683.
10. Scheuplein RJ, Blank IH, Brauner GJ, Macfarlane DJ (1969) Percutaneous absorption of steroids. J Investig Dermatol 52: 63-70.
11. Flynn G (1990) Physiochemical determinants of skin absorption. In: Gerrity TR (ed.). Principles of route-to-route extrapolation for risk assessment. Elsevier, New York, USA, pp. 93-127.
12. Guy RH, Potts RO (1992) Structure-permeability relationships in percutaneous penetration. J Pharm Sci 81: 603-604.
13. Potts RO, Guy RH (1992) Predicting skin permeability. Pharmaceutical Research 9: 663-669.
14. Tayar NE, Tsai RS, Testa B, Carrupt PA, Hansch C, et al. (1991) Percutaneous penetration of drugs: a quantitative structure-permeability relationship study. J Pharm Sci 80: 744-749.
15. Cronin MT, Dearden JC, Moss GP, Murray-Dickson G (1999) Investigation of the mechanism of flux across human skin in vitro by quantitative structure-permeability relationships. Eur J Pharm Sci 7: 325-330.
16. Abraham MH, Chadha HS, Mitchell RC (1995) The factors that influence skin penetration of solutes. J Pharm Pharmacol 47: 8-16.
17. Abraham MH, Martins F, Mitchell RC (1997) Algorithms for skin permeability using hydrogen bond descriptors: the problem of steroids. J Pharm Pharmacol 49: 858-865.
18. Lien EJ, Gaot H (1995) QSAR analysis of skin permeability of various drugs in man as compared to in vivo and in vitro studies in rodents. Pharm Res 12: 583-587.
19. Magee PS (1998) Some novel approaches to modelling transdermal penetration and reactivity with epidermal proteins: Comparative QSAR. Taylor & Francis, London, pp: 137-168.
20. Potts RO, Guy RH (1995) A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. Pharm Res 12: 1628-1633.
21. Pugh WJ, Roberts MS, Hadgraft J (1996) Epidermal permeability-penetrant structure relationships: the effect of hydrogen bonding interactions and molecular size on diffusion across the stratum corneum. Int J Pharm 138: 149-165.
22. Pugh WJ, Degim IT, Hadgraft J (2000) Epidermal permeability-penetrant structure relationships: QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. Int J Pharm 197: 203-211.
23. Raevsky OA, Fetisov VI, Trepalina EP, McFarland JW, Schaper KJ (2000) Quantitative estimation of drug absorption in humans for passively transported compounds on the basis of their physico-chemical parameters. Quantitative Structure-Activity Relationships 19: 366-374.
24. Ghasemi JB, Rofouei MK, Salahinejad M (2011) A quantitative structure-property relationships study of the stability constant of crown ethers by molecular modelling: new descriptors for lariat effect. J Incl Phenom Macrocycl Chem 70: 37-47.

25. Ghasemi J, Saaidpour S (2008) QSPR modeling of stability constants of diverse 15-crown-5 ethers complexes using best multiple linear regression. J Incl Phenom Macrocycl Chem 60: 339-351.

26. Baba H, Takahara JI, Mamitsuka H (2015) In silico predictions of human skin permeability using nonlinear quantitative structure-property relationship models. Pharm Res 2: 2360-2371.

27. Puzyn T, Mostrag-Szlichtyng A, Gajewicz A, Skrzynski M, Worth AP (2011) Investigating the influence of data splitting on the predictive ability of QSAR/QSPR models. Struct Chem 22: 795-804.

28. Duran A, Pastor M (2011) An advanced tool for computing and handling GRid-INdependent. Descriptors User Manual Version 1.

29. Goodford PJ (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. J Med Chem 28: 849-857.

30. Duran A, Martinez GC, Pastor M (2008) Development and validation of AMANDA, a new algorithm for selecting highly relevant regions in molecular interaction fields. J Chem Inf Model 48: 1813-1823.

31. Ghasemi J, Tavakoli H (2015) Improvement of the prediction power of the CoMFA and CoMSIA models on histamine H3 antagonists by different variable selection methods. Sci Pharm 80: 547-566.

32. Scheerlinck K, Debaets B, Stefanov I, Fievez V (2010) Subset selection from multi-experiment data sets with application to milk fatty acid profiles. Comput Electron Agric 73: 200-212.

33. Melanie M (1999) An introduction to genetic algorithms. A Bradford Book. The MIT Press, London, England, pp: 62-75.

34. Geladi P, Kowalski BR (1986) Partial least-squares regression: a tutorial. Anal Chim Acta 185: 1-7.

35. Wold S, Sjostrom M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometr Intell Lab Syst 58: 109-130.

36. Cratteri P, Romanelli MN, Cruciani G, Bonaccini C, Melani F (2004) GRIND-derived pharmacophore model for a series of α-tropanyl derivative ligands of the sigma-2 receptor. J Comput Aided Mol Des 18: 361-374.

37. Ghasemi JB, Hooshmand S (2013) 3D-QSAR, docking and molecular dynamics for factor Xa inhibitors as anticoagulant agents. Mol Simul 39: 453-471.

38. Kabeya LM, da Silva CH, Kanashiro A, Campos JM, Azzolini AE, et al. (2008) Inhibition of immune complex-mediated neutrophil oxidative metabolism: a pharmacophore model for 3-phenylcoumarin derivatives using GRIND-based 3D-QSAR and 2D-QSAR procedures. Eur J Med Chem 43: 996-1007.

39. Ragno R, Simeoni S, Rotili D, Caroli A, Botta G, et al. (2008) Class II-selective histone deacetylase inhibitors. Part 2: alignment-independent GRIND 3-D QSAR, homology and docking studies. Eur J Med Chem 43: 621-632.

40. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemometr Intell Lab Syst 2: 37-52.

41. Efron B (1983) Estimating the error rate of a prediction rule: improvement on cross-validation. J Am Stat Assoc 78: 316-331.

42. Geisser S (1975) The predictive sample reuse method with applications. J Am Stat Assoc 70: 320-328.

43. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20: 273-297.

44. Vapnik VN (1988) Statistical learning theory. John Wiley & Sons Inc., New York, NY, USA.

45. Afantitis A, Melagraki G, Sarimveis H, Igglessi-Markopoulou O, Kollias G (2009) A novel QSAR model for predicting the inhibition of CXCR3 receptor by 4-N-aryl-[1, 4] diazepane ureas. Eur J Med Chem 44: 877-884.

46. Darnag R, Mazouz EM, Schmitzer A, Villemin D, Jarid A (2010) Support vector machines: development of QSAR models for predicting anti-HIV-1 activity of TIBO derivatives. Eur J Med Chem 45: 1590-1597.

47. Stanforth RW, Kolossov E, Mirkin B (2007) A measure of domain of applicability for QSAR modelling based on intelligent K-means clustering. QSAR Comb Sci 26: 837-844.

48. Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, et al. (2008) Critical assessment of QSAR models of environmental toxicity against Tetrahymena pyriformis: focusing on applicability domain and overfitting by variable selection. J Chem Inf Model 48: 1733-1746.

49. Tropsha A, Golbraikh A (2010) Predictive quantitative structure-activity relationships modeling: development and validation of QSAR models. In: Faulon JL, Bender A (eds.). Handbook of chemoinformatics algorithms. Chapman and Hall, CRC Press, USA, pp: 223-244.

50. Pham H, Gonzalez-Alvarez I, Bermejo M, Mangas Sanjuan V, Centelles I, et al. (2011) In Silico Prediction of Caco-2 Cell Permeability by a Classification QSAR Approach. Mol Inform 30: 376-385.