How to Generate Reliable and Predictive CoMFA Models

L. Zhang¹, K.-C. Tsai², L. Du¹, H. Fang¹, M. Li*¹ and W. Xu*¹

¹Department of Medicinal Chemistry, School of Pharmacy, Shandong University, Jinan, Shandong 250012, China
²The Genomics Research Center, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 115, Taiwan

Abstract: Comparative Molecular Field Analysis (CoMFA) is a mainstream and down-to-earth 3D QSAR technique in the coverage of drug discovery and development. Even though CoMFA is remarkable for high predictive capacity, the intrinsic data-dependent characteristic still makes this methodology certainly be handicapped by noise. It’s well known that the default settings in CoMFA can bring about predictive QSAR models, in the meanwhile optimized parameters was proven to provide more predictive results. Accordingly, so far numerous endeavors have been accomplished to ameliorate the CoMFA model’s robustness and predictive accuracy by considering various factors, including molecular conformation and alignment, field descriptors and grid spacing. Herein, we would like to make a comprehensive survey of the conceivable descriptors and their contribution to the CoMFA model’s predictive ability.

Keywords: CoMFA, conformation, alignment, fields, grid spacing.

1. INTRODUCTION

Quantitative structure-activity relationship (QSAR) studies generally perform a crucial role in drug discovery and design as a ligand-based approach [1]. Such approaches are explicitly judgmental to provide not only the reliable prediction of specific properties of new compounds, but also the help to elucidate the possible molecular mechanism of the receptor-ligand interactions, in case that the experimental NMR or crystal structure of the target protein is unavailable [2]. As one of the most sought-after QSAR methods, Comparative Molecular Field Analysis (CoMFA) recruits interactive graphics and statistical techniques for correlating several molecular features, such as steric and electrostatic properties with their biological activities [3]. It needs to be noted that over the past few decades, CoMFA has become tremendously prevalent in regions of both industrial and academic research regarding QSAR studies. A Scifinder Scholar survey with the term ‘CoMFA’ indicated about 160 publications in 2009. This result compares with only about 50 publications in 1995 (Fig. 1). These CoMFA applications in drug design have been comprehensively summarized in several excellent book chapters as well as review articles [4-6].

In CoMFA philosophy, the biological properties of molecules are correlated with steric and electrostatic potential energies in terms of Lennard-Jones and Coulombic potentials, respectively. The steric and electrostatic potential energies are calculated by a probe atom, located at each vertex of a spaced lattice, in which a series of molecules are embedded. The performance of the standard CoMFA procedure requires the specification of both conformations and alignments of molecules. Despite their popularity, CoMFA analyses can be highly sensitive to the parameters used in QSAR modeling, including the setting of steric fields, molecular alignments, and grid spacing and dimensions. Because of these variances, there have been interests in finding ways to enhance QSAR quality and in building robust CoMFA models.

The common steps for CoMFA modeling (Fig. 2) include:

1. Active molecules are placed in a three-dimensional grid (2Å spacing) encompassing all of the molecules.
2. At each grid point, steric energy (Lennard-Jones potential) and electrostatic energy are measured for each molecule by a probe atom ($sp^3$-hybridized carbon with +1 charge).
3. To minimize domination by large steric and electrostatic energies, all energies that exceed a specified value (default 30 kcal/mol) are set to the cutoff value.
4. CoMFA uses a partial least-squares (PLS) analysis to predict activity from energy values at the grid points.

Based on these procedures, hitherto we would like to review the up-to-date literatures for enhancing the quality of CoMFA modeling herein.

CoMFA Settings

It is a strenuous job for modelers to determine which settings, or combinations of settings, are suitable for their data sets. Notwithstanding the fact that a number of settings are provided, most CoMFA users still rely on the default parameters (Fig. 3). Hence, the influences of adjusting one or more CoMFA settings such as steric molecular field settings, grid distances and cut off values have been well explored by several studies [7-10].

Karlen and coworkers produced a total of 6120 CoMFA models by settings optimization to evaluate the possibility of improving the predictive ability [11]. This effort was evaluated by nine different data sets, and the optimal models shared ordinary feature of steric molecular field settings, grid distances and cut off values have been well explored by several studies [7-10].

Molecular Conformation

Cramer and his coworkers have annotated that “active conformation” and “alignment rules” are major would-be dilemmas for
CoMFA [12]. Data sets for CoMFA examination should manoeuvre through the same mechanism of action, as well as have a common pharmacophore even if the molecular skeleton is different [4, 5, 13]. Three dimensional structures of the selected compounds are mandatory for a CoMFA modeling, along with the bioactive shapes as prerequisite for a credible model. It’s pretty recognized that construction of bioactive conformation for a set of molecules is arduously realizable by way of crystallographic studies. Nevertheless, diverse computational approaches, including docking, molecular dynamic simulation and conformational sampling, are prosperously ready for use to predict the bioactive conformation.

Provided that the holo-form crystal structure for target-ligand complex is attainable and the data set compounds have an analogous scaffold with the ligand, this cocryostalized ligand can be of course applied as a template for establishing the framework of the selected data set. This method is straightforward but common-used in the 3D-QSAR resolution [14-18]. It needs to be remarked that this method can be exclusively plausible for a bunch of rigid and structurally similar molecules. Notwithstanding, for a collection of highly structurally flexible or diverse molecules, this method is too much simple to be persuasive.

A common used approach for predicting the bioactive configuration is molecular docking. In the docking procedure, a ligand was positioned to the functional site of target for determining the binding mode of the complex as well as generating the active conformation of the ligand. So far a number of predictive CoMFA models were successfully generated based on docking conformations, such as Chretien’s acetylcholinesterase (AChE) model [19], Bharatam’s glycogen synthase kinase-3 (GSK-3) model [20], Qiu’s c-Jun N-terminal kinase-1 (JNK-1) model [21], and Yao’s vascular endothelial growth factor receptor tyrosine kinase-2 (VEGFR-2) model [22].

Molecular dynamic simulation of the ligand-target complex can still provide a precise conformation. There are numerous successful examples for CoMFA modeling based on MD simulation, for ex-
ample. Shang and coworkers built the initial structures of inhibitors by MD simulation of enzyme-substrate complexes for their CoMFA study [23]. Nevertheless, such accurate calculation is still too time-consuming and resource-dependent to be applied in QSAR modeling.

Bioactive conformational analysis has to be in silico simulated in case that the receptor experiential structure is not available. A systematic search [24-26] or annealing process [27-29] can be usually utilized to generate the low energy conformation for a template molecule; subsequently the remaining molecules in the dataset can be constructed based on the above-mentioned reference conformation. Active analog approach generates the possible conformations by conformational analysis and selects the best ones that satisfy the interatomic distances in a working hypothesis [30-32]. Local minimization method determines active conformations through systematic conformational searches followed by minimization on the rigid rotor search surface [33]. Nonetheless, this procedure can exclusively lead to the minimized conformations, even if the bioactive conformation is not equal to the lowest energy.

A “similarity” problem must be figured out is that if compounds in the test set have significant conformational diversity compared with the training set molecules (contain certain features in a region not explored by the training set), the CoMFA model generated based on the training set can’t accurately predict activities of the test set [34].

Molecular Alignment

Molecular superposition plays a decisive role in CoMFA analysis, since the relative interaction energies depend strongly on relative molecular positions. Flexible groups and conformation diversity present dilemmas to molecular alignment. In this case, a low quality alignment can result in unexpected input noise. Pseudoreceptor modeling [35-40] and 4D QSAR analysis [41-43] show high performance because they have the advantage of reducing input noise and decreasing risk of overfitting. As to CoMFA, various methods were developed aiming to seek a strict or a “perfect” molecular overlay.

A receptor-based alignment approach can be achieved by the molecular docking method. In this procedure, similar molecules were located to the same region of the active site, and subsequently, an automatic alignment was generated. Using alignment derived from this receptor-based approach can generate highly predictive CoMFA models, and even shows superiority to the ligand-based method [40, 44-46].

The pharmacophore modeling is another automatic molecular alignment generation approach. Unlike molecular docking, this method is independent on the receptor structure. It determines the pharmacophore features of the ligands and uses these features to align the molecular structures. Since the problem of conformational diversity and flexibility is solved in some distance, it is widely applied in construction of CoMFA models [47-50].

Force fields can be used to align molecules and the field fit algorithm is used to calculate field values of the grid points. Lattice points with certain field value magnitude are employed for molecular overlapping. Dove and coworkers compared three different alignment rules, and found that the weighted field fit alignment rule gave the best model [51]. It is concluded that the weighted field fit method reduced the risk of producing artificial redundancy of the structures and ignoring entropy contributions to the free energy of binding.

Even though the automatic alignment rules are widely practiced in the CoMFA approach with high performance, the manual alignment method is a competent alternative for yielding high predictive models as well. Tervo and coworker derived CoMFA models based on two alignment rules, docking and manual methods, for a set of flexible molecules [52]. Interestingly, the results suggested that model with better predictive quality was constructed on the basis of manual alignment rule.

The brandnew topomer CoMFA method introduced since 1998 can break the input structures into fragments and removes any core fragment structurally common to the entire series [53]. The field values are calculated for the left fragments. It should be spotlighted that this novel fragment-based alignment method is proved to be promising CoMFA construction approach for both QSAR experts and newbies within few minutes, and results are generally equivalent to canonical CoMFA analysis [53, 54].

CoMFA Fields

In CoMFA analysis, a collection of structurally aligned molecules are represented in terms of property fields, which are evaluated at a probe atom and each molecule at regularly spaced intervals on a grid. The acquiescent CoMFA fields, steric and electrostatic fields, are calculated by Lennard-Jones potential (eq 1) and Columbic potential (eq 2), respectively.

\[
E_{\text{vdw}} = \sum_{i=1}^{n} \left( A_{ij} \frac{r_{ij}^{-12}}{C_{ij} r_{ij}^6} \right)
\]

\[
E_C = \sum_{i=1}^{n} \frac{q_i q_j}{D r_{ij}^6}
\]

where \(E_{\text{vdw}}\) is van der Waals interaction energy, \(r_{ij}\) is distance between atom \(i\) of the molecule and the grid point \(j\) where the probe atom is located, \(A_{ij}\) and \(C_{ij}\) are constants depending on the van der Waals, radii of the corresponding atoms. \(E_C\) is coulomb interaction energy, \(q_i\) is partial charge of atom \(i\) of the molecule, \(q_j\) is charge of the probe atom, \(D\) is dielectric constant.

Both potential functions are remarkably steep nearby the van der Waals surface of the molecules, thus resulting in rapid changes in surface descriptions (Fig. 4). The steric fields designate Van der Waals interactions between the molecule and its receptor. The standard CoMFA method functions the Lennard Jones 6-12 potential, which is characterized by a highly steep enhancement in energy at short distances to calculate the steric interaction between a probe atom and a molecule. So far different methods were applied to calculate steric field, their contribution to QSAR quality were evaluated as well [55, 56]. The indicator method assigned an energy 0 to steric potentials which fall below the cutoff value, and assigned a nominal energy (equal to the cutoff) to potentials falling at or above the cutoff [34]. The parabolic approach made the magnitude of the calculated potential at each lattice point be squared [11]. Gaussian function which has a slower and smoother decrease was facilitated in CoMSIA method for calculating molecular field potentials.

To prevent unjustified large parametric variance, steric energy truncated at lower values was well documented because of the steep increase of the steric field contribution at lattice points close to the molecule. For example, Kim and Martin truncated the steric energy to 4.0 [58], in the meanwhile Klebe and coworkers set two different energy truncation of 30 and 5 for evaluating the effect of energy truncation [59].

In 2008, Sorich and coworkers published their studies for investigating the guidance of steric field settings on the predictive performance of CoMFA [7]. In this case, 3D QSAR models based on Lennard-Jones, indicator, parabolic and Gaussian steric fields were sensibly compared using 28 datasets. The results demonstrated that the preformance of Lennard-Jones and indicator fields that have a steep value decrease was better than of Gaussian type, which have a smoother decrease.
As one predominant effecter on CoMFA generation, electrostatic field is typically brought about by calculating the Coulomb potential between a probe and the molecule. The empirical partial charge methods, including Gasteiger-Marsili, Gasteiger-Hückel, and MMFF94, are comprehensively used in released CoMFA studies. Developed on the concept of equalization of electronegativities, these empirical approaches are simple-handled and quick-witted in assigning the partial charges. The common-used semiempirical methods, including the modified neglect of differential overlap (MNDO), Austin model 1 (AM1) and parametric model 3 (PM3), are a tradeoff between the empirical and \textit{ab initio} quantum chemical approaches (HF/STO-3G, HF/3-21G* and HF/6-31G*) in terms of precision and computational time. The \textit{ab initio} approaches are of high meticulousness but low computational efficiency. Contrarily, the empirical methods have fast speed but relative low accuracy. The semi-empirical techniques have enhancement in speed over the \textit{ab initio} methods. Nevertheless, they did not have significant improvement on accuracy when compared with the empirical methods.

Hitherto a number of studies were accomplished to evaluate the importance of partial charges on CoMFA model’s predictive quality. [60-63] For example, Welsh and coworkers performed CoMFA on a set of human immunodeficiency type 1 (HIV-1) protease inhibitors and different charge assignment schemes (Discover CVFF, Gasteiger-Marsili and AM1-ESP) were evaluated [64]. The best model was constructed using AM1-ESP partial charges.

Recently, Sorich and coworker examined the contribution of partial charge calculation methods to the predictive ability of the generated CoMFA models [65]. In their study, Gasteiger, Gasteiger-Huckel, MMFF94, AM1, MNDO and PM3 charges were assigned to 30 data sets. The authors found that semi-empirical charge calculation methods suggested for the most predictive models, MMFF94 was also a good alternative for its predictive ability (not significantly worse than the semiempirical methods) and fast calculation speed.

We also did a study for comparing twelve semi-empirical and empirical charge-assigning methods, including AM1, AM1-BCC, CFF, Del-Re, Formal, Gasteiger, Gasteiger-Hückel, Hückel, MMFF, PRODRG, Pullman and VC2003, on prediction accuracy in CoMFA and CoMSIA studies by using several benchmark datasets [66]. In general, the semi-empirical charges, such as AM1 and AM1-BCC, provided higher predictive CoMFA and CoMSIA models than the Gasteiger and Gasteiger-Hückel charges which are commonly used in QSAR studies. Interesting, the CFF partial charge was found to obtain the most predictive CoMFA and CoMSIA models. As an empirical charge-assigning method with a short computing time, the CFF charge offer advantages over the other eleven semi-empirical and empirical charges in performing the most accurate electrostatic potential calculations for CoMFA and CoMSIA studies. These results presented should help the selection of electrostatic potential models in CoMFA and CoMSIA studies.

**Grid Spacing**

Kroemer and coworker elucidated that the appearance of the molecular structures should be precisely characterized in the lattice, whereas the degree of differentiation should not be excessively high [33]. For that reason, an exceptional grid is constrained to discriminate atoms of different molecules and then formulate the corresponding values into the descriptor matrix.

The influence of grid spacing has been thoroughly evaluated in foregoing surveys. Consequently, these results demonstrated that the domination of lattice location and size is not appreciable because of limited datasets [67, 68]. Tropsha and coworker presented that changing the dimensions of the region can lead to an unreasonable predictive ability of CoMFA model [69]. They mentioned that this characteristic should be well taken into account when operating CoMFA. Likewise, Sorich and coworkers substantiated that there is a statistically considerable function of grid spacing on predictability for convinced steric field settings [7]. They also manifested that lattice density has a somewhat subsidiary influence on the predictive capacity of Gaussian steric fields, whereas such parameters are able to manipulate the CoMFA quality based on Lennard-Jones, indicator and parabolic steric fields.

**Other Descriptors**

Hydrophobic property was frequently thought about in CoMFA studies, for enlightening the CoMFA models’ robustness and predictive capability [70-72]. Wiese and coworkers established more than 350 CoMFA models and then benchmarked them using steric, electrostatic and hydrophobic fields alone and in combination [73]. These consequences indicated that hydrophobic fields could boost the correlative and predictive power in all cases. In the following year, they compared the model quality by 3D (HINT hydrophobic field) and logP (HINT and ClogP values) presentations of hydrophobicity, as well as evaluated the likeness between standard CoMFA and hydrophobic fields [74]. They further exhibited that ClogP is prior to logP when deriving hydrophobic fields for the generation of CoMFA model. This thought-provoking evidence uncovered that hydrophobic properties can perform the indicative role in describing the molecule. Even so, in the event that the hydrophobic properties have no correlation with the activity, these properties may present negative contribution to the QSAR quality.

So far a number of examinations showed that consideration of H-bond descriptors can improve the quality of the CoMFA model certainly [75, 76]. Xu and coworkers carried out CoMFA models on a series of gamma-hydroxy butenolide endothelin antagonists in the presence of additional H-bond fields [77]. The results confirmed that H-bond fields significantly improved the quality of the derived model. Moreover, Pan and coworkers emphasized the influence of H-bond fields on improving the CoMFA model quality based on a set of protein tyrosine phosphatase 1B (PTP1B) inhibitors [78].

The guidances of frontier orbital (HOMO and LUMO) energies as additional physical descriptors were also evaluated in diverse
QSAR studies. In result, the introduction of HOMO and LUMO are proven to be suggestive for the acquired CoMFA models [79-82]. The incorporation of ligand-receptor energies to the CoMFA model can contribute to the model quality as well [83].

In CoMFA modeling, a $sp^3$ carbon with $+1.0$ charges is generally applied as the default probe atom. However, the selection of probe atoms can involve in the CoMFA generation. Hannongbua and coworker examined the impacts of different probe atoms ($Csp^3(+-1)$, $Os(-1)$ and $H(+1)$) on the quality of the generated CoMFA models. Their results elucidated that a combination of these three probes led to the most predictive model [84].

**Statistical Analysis**

Partial least squares (PLS), developed in 1986, is a prevailing statistical algorithm for deriving linear relationships among columns of data [85-88]. Theoretically, PLS, a regression function, looks for linear correlation of column variance in target properties with variations in explanatory properties to minimize the sum of squares of deviations. It is like a factor analysis of the explanatory properties in which the object is to maximize alignment with the target property values rather than with the Cartesian or other axes. For this feature, PLS analysis is sometimes compared to principal component regression analysis (PCA) in its derivation of vectors from the Y and X blocks. PCA, introduced by Karl Pearson, involves a mathematical procedure that transforms a number of possibly correlated variables into the smaller number of uncorrelated variables [89]. Unlike PCA, which only considers the influence of the input data array, PLS possesses both the input and the output data matrices into consideration.

Another critical algorithm, SAMPLS, created by Bruce Bush, tremendously accelerates the cross-validation procedure [90]. In SAMPLS, the latent variables are derived from the $n \times n$ covariance matrix (Fig. 5). Validation methods such as cross-validation (including leave-one-out [12], leave-many-out [91-93] and leave-group-out [94]) and bootstrapping [95-97] have employed to examine the robustness of the generated models.

The quality of the resulting QSAR models can be judged by statistical means such as $r^2$ (the fraction of explained variance, eq. 3) for the test set, and by $q^2$ (the cross-validated or predictive $r^2$, eq. 4) for the training set. The fraction of explained variance, $r^2$, measures the Q SAR model’s ability to interpret the variance in the data; in other words, it estimates the goodness-of-fit of the regression model derived from the training set. The predictive $r^2$, or $q^2$, refers to the internal robustness of the QSAR model. Model quality estimation are procured either by using cross-validation procedure (internal) or by predicting external compounds (previously not used in the model).

$$
r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}
$$

$$
q^2 = 1 - \frac{PRESS}{SD} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}
$$

Where $Y_i$ is an actual value, $\hat{Y}_i$ is a predictive value, $\bar{Y}$ is the average value of predicted values, $Y_m$ is the average value of observed activities. $PRESS$ = predictive sum of squares, $SD$ = sum of squared deviations.

Recently Tropsha and coworkers introduced a new validation criterion for robust QSAR models (eq. 5-8) [98, 99]. They consider a QSAR model predictive, if the following conditions are satisfied:

From $u = \mathbf{Xt}$ ($k = \text{constants}$ $k1, k2 \ldots k_j$; $j = \text{number of PLS components}$) follows:

$$
\hat{Y}_i = a_1S_{i1} + a_2S_{i2} + a_3S_{i3} + \ldots + b_1E_{ij} + b_2E_{j2} + \ldots + b_nE_{jn}
$$

where $\mathbf{X}$ and $\mathbf{Xt}$ are the input data array, $\mathbf{P}$ possesses both the input and the output variables [89]. Unlike PCA, which only considers the influence of the input data array, PLS possesses both the input and the output data matrices into consideration. Another critical algorithm, SAMPLS, created by Bruce Bush, tremendously accelerates the cross-validation procedure [90]. In SAMPLS, the latent variables are derived from the $n \times n$ covariance matrix (Fig. 5).

Validation methods such as cross-validation (including leave-one-out [12], leave-many-out [91-93] and leave-group-out [94]) and bootstrapping [95-97] have employed to examine the robustness of the generated models.

The quality of the resulting QSAR models can be judged by statistical means such as $r^2$ (the fraction of explained variance, eq. 3) for the test set, and by $q^2$ (the cross-validated or predictive $r^2$, eq. 4) for the training set. The fraction of explained variance, $r^2$, measures the Q SAR model’s ability to interpret the variance in the data; in other words, it estimates the goodness-of-fit of the regression model derived from the training set. The predictive $r^2$, or $q^2$, refers to the internal robustness of the QSAR model. Model quality estimation are procured either by using cross-validation procedure (internal) or by predicting external compounds (previously not used in the model).

$$
r^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}
$$

$$
q^2 = 1 - \frac{PRESS}{SD} = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}
$$

Where $Y_i$ is an actual value, $\hat{Y}_i$ is a predictive value, $\bar{Y}$ is the average value of predicted values, $Y_m$ is the average value of observed activities. $PRESS$ = predictive sum of squares, $SD$ = sum of squared deviations.

Recently Tropsha and coworkers introduced a new validation criterion for robust QSAR models (eq. 5-8) [98, 99]. They consider a QSAR model predictive, if the following conditions are satisfied:
used for building the model. Models with high internal $q^2$ but low external predictive abilities can be picked out by using these equations that were verified by 160 QSAR models.

CONCLUSION

The CoMFA technique has been developed for more than one couple of decades. Thus far a great number of CoMFA studies were performed based on this state-of-the-art approach. Scientists have also contributed everlasting and booming endeavors to improve the predictive quality of the CoMFA model. Herein, the practicable CoMFA descriptors, including molecular conformation, structural alignment, molecular fields, grid spacing and additional physical chemical properties, were well presented as a tutorial review to provide possible guidance to the further CoMFA studies. Among these crucial determinants, bioactive conformation and molecular superposition engage an essential portrayal in the CoMFA procedure, while different combination of fields and physical chemical properties results in diverse predictable levels. High predictive models can also be realized by adjusting settings, such as energy cutoff values, lattice size and probe types.

In sum, suggestions for future CoMFA studies are outlined below.

1. The initial geometries of the molecules should be in bioactive or theoretical active framework;
2. Different charge methods should be carefully considered to establish a muscular CoMFA model;
3. A reasonable molecular alignment is mandatory for a trustworthy CoMFA model;
4. Cut-off values are needed both for the steric and electrostatic energy calculation and for the PLS analysis to reduce unwanted variance;
5. Other descriptors, such as ClogP, can substantially improve the reliability of the CoMFA model. In the absence of statistic significance in CoMFA generation, those descriptors can be taken into consideration;
6. Different probe atoms could be attentively considered to ameliorate the credibility of CoMFA model;
7. The lattice location and size should be unanimously deliberated.

ACKNOWLEDGEMENTS

The present work was financed by grants from the PhD Programs Foundation of Ministry of Education of China (No. 20090131120080), the Doctoral Fund of Shandong Province (No. BS2009SW011), Fok Ying Tung Education Foundation (No. 122036), National Natural Science Foundation of China (No. 30901836 and 81001362), Shandong Natural Science Foundation (No. JQ201019) and Independent Innovation Foundation of Shandong University, IIFSDU (No. 2010JQ005).

ABBREVIATIONS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CoMFA</td>
<td>comparative molecular field analysis</td>
</tr>
<tr>
<td>QSAR</td>
<td>quantitative structure-activity relationship</td>
</tr>
<tr>
<td>PLS</td>
<td>partial least-squares</td>
</tr>
<tr>
<td>PCA</td>
<td>principal component analysis</td>
</tr>
<tr>
<td>MD</td>
<td>molecular dynamic</td>
</tr>
<tr>
<td>3D-QSAR</td>
<td>three dimensional quantitative structure-activity relationship</td>
</tr>
<tr>
<td>CoMSIA</td>
<td>comparative molecular similarity indices analysis</td>
</tr>
<tr>
<td>AChE</td>
<td>acetylcholinesterase</td>
</tr>
<tr>
<td>GSK-3</td>
<td>glycogen synthase kinase-3</td>
</tr>
<tr>
<td>JNK-1</td>
<td>c-Jun N-terminal kinase-1</td>
</tr>
<tr>
<td>VEGFR-2</td>
<td>vascular endothelial growth factor receptor tyrosine kinase-2</td>
</tr>
<tr>
<td>MND0</td>
<td>modified neglect of differential overlap</td>
</tr>
<tr>
<td>AM1</td>
<td>Austin model 1</td>
</tr>
<tr>
<td>PM3</td>
<td>parametric model 3</td>
</tr>
<tr>
<td>HIV-1</td>
<td>human immunodeficiency type 1</td>
</tr>
<tr>
<td>PTP1B</td>
<td>protein tyrosine phosphatase 1B</td>
</tr>
<tr>
<td>SAMPLES</td>
<td>sample-distance partial least squares</td>
</tr>
</tbody>
</table>

REFERENCES


