The effect of various atomic partial charge schemes to elucidate consensus activity-correlating molecular regions: a test case of diverse QSAR models

Sivakumar Prasanth Kumar\textsuperscript{a}, Prakash C. Jha\textsuperscript{b}, Yogesh T. Jasrai\textsuperscript{**} and Himanshu A. Pandya\textsuperscript{a}

\textsuperscript{a}Department of Bioinformatics, Applied Botany Centre (ABC), University School of Sciences, Gujarat University, Ahmedabad 380 009, India; \textsuperscript{b}School of Chemical Sciences, Central University of Gujarat, Sector-30, Gandhinagar 382030, India

Communicated by Ramaswamy H. Sarma

(Received 14 January 2015; accepted 21 April 2015)

The estimation of atomic partial charges of the small molecules to calculate molecular interaction fields (MIFs) is an important process in field-based quantitative structure–activity relationship (QSAR). Several studies showed the influence of partial charge schemes that drastically affects the prediction accuracy of the QSAR model and focused on the selection of appropriate charge models that provide highest cross-validated correlation coefficient ($r^2_{cv}$ or $q^2$) to explain the variation in chemical structures against biological endpoints. This study shift this focus in a direction to understand the molecular regions deemed to explain SAR in various charge models and recognize a consensus picture of activity-correlating molecular regions. We selected eleven diverse dataset and developed MIF-based QSAR models using various charge schemes including Gasteiger-Marsili, Del Re, Merck Molecular Force Field, Hückel, Gasteiger–Hückel, and Pullman. The generalized resultant QSAR models were then compared with Open3DQSAR model to interpret the MIF descriptors decisively. We suggest the regions of activity contribution or optimization can be effectively determined by studying various charge-based models to understand SAR precisely.

Keywords: molecular interaction fields; field-based QSAR; atomic partial charge schemes; consensus descriptors; COX-2 inhibitors

1. Introduction

Three-dimensional quantitative structure–activity relationship (3D-QSAR) is an important computational approach in ligand-based drug discovery and design (Lill, 2007). It explores the relationship between properties of chemical substances and their biological endpoints to develop a statistical model, and it is based on the fundamental principle that the difference in structural properties is responsible for observed variations in the biological response of the molecules (Verma, Khedkar, & Coutinho, 2010). The concept that molecules interact non-covalently through their electronic properties such as electrostatic and van der Waals forces, revolutionized the perspectives of medicinal chemistry arena and essentially made way to pioneered research on molecular interaction fields (MIFs) (Cramer, Patterson, & Bunce, 1988; Wermuth, 1996). A MIF explains the spatial variation of the interaction energy between a molecular target (a macromolecule or a chemical substance) and a chosen probe (Wade, 1993). Particularly, the progress in the MIFs studies was attributed to the development of three popular QSAR methods, Comparative Molecular Field Analysis (CoMFA) (Cruciani, 2006), Comparative Molecular Similarity Indices Analysis (CoMSIA) (Klebe, Abraham, & Mietzner, 1994), and the GRID force field (Goodford, 1985). The terminology MIF has been used synonymously with the terms, molecular field analysis (MFA), field-based QSAR and GRID-based QSAR regardless of the QSAR softwares, in the scientific community (Consonni & Todeschini, 2010).

A brief introduction of various molecular field technologies is given here. The standard CoMFA method involves the placement of aligned small molecules centrally in a grid box or lattice grid with a regular spacing or interval. Further, the steric and electrostatic potential energies were computed using a probe at all intersections of the lattice and subsequently, utilized as independent variables to correlate with biological endpoints using partial least square (PLS) statistical technique (Cramer et al., 1988; Cruciani, 2006). Various breakthroughs in CoMFA method have been witnessed since decades and discussed elaborately in the literature (Consonni & Todeschini, 2010; Verma et al., 2010; Wade, 1993). To overcome some of the limitations in CoMFA method, CoMSIA was developed which calculates molecular similarity indices from Steric and Electrostatic A Liability (SEAL) similarity fields (Kubinyi, Hamprecht, & Mietzner, 1998), to generate steric,
The implementation of bell-shaped Gaussian functions of SEAL fields that provide good approximations of Lennard-Jones and Coulomb potentials in contrast to its steep nature in CoMFA, made CoMSIA a famous QSAR tool (Klebe, Mietzner, & Weber, 1994). The GRID program based on GRID force field emerged as an alternative to the CoMFA method for estimating interaction fields and possesses an advantage of employing a comparatively smoother 6–4 Lennard-Jones potential than 6–12 form in CoMFA (Goodford, 1985). Moreover, this program applied a new concept of variable selection and reduction called GOLPE (Generating Optimal Linear PLS Estimations) approach to recognize highly informative variables from the descriptor pool, and hence, this technique is collectively called GRID/GOLPE approach (Baroni et al., 1993). The application of molecular field extrema as descriptors for QSAR modeling was introduced in FieldAlign program to overcome grid restrictions and sample large numbers of descriptors (Cheeseright, Mackey, Rose, & Vinter, 2006). These descriptors were derived from the eXtended Electron Distribution force field which treats the electronic distribution as multipole in accordance with the quantum orbital descriptors rather than conventional atom-centered charges and utilizes Gaussian functions (Vinter, 1994).

The assumption of a nonlinear relationship between the dependent and independent variables in $k$-nearest neighbor ($k$NN) method benefitted with other stochastic variable selection methods such as simulated annealing and evolutionary algorithms, allowed the development of QSAR models without the need to obtain a single linear equation. In this method, $k$NNs of any molecule are acquired by computing the distance between the descriptors chosen from numerous variable selection methods (Ajmani, Jadhav, & Kulkarni, 2006). Field-based QSAR (also known as force field for CoMFA-like models and Gaussian for CoMSIA-like models) is a tool from Schrödinger, LLC which works on the basic principles of CoMFA/CoMSIA and utilizes a specific set of parameters including the derivation of steric potentials from the OPLS (Optimized Potentials for Liquid Simulations)-2005 force field, hydrophobic properties from CLOGP fragmental methods, and H bond acceptor (HBA) and donor fields from its own Phase pharmacophore module (Field-Based QSAR (version 2013-3), 2013). A free, open-source chemometric tool designed to explore pharmacophore by high-throughput MIFs analysis and equipped with automations through scriptable interface, flexibility and interoperability with external programs as well as algorithm parallelization for variable selection and model validation to achieve high performance, make Open3DQSAR a powerful tool in pharmacophore assessment and ligand-based drug design (Tosco & Balle, 2011).

The development of non-steroidal anti-inflammatory drugs that exhibit anti-inflammatory roles primarily via inhibition of cyclooxygenase (COX) has been widely explored in the treatment of pain and inflammation (Dannhardt & Kiefer, 2001). When induced, COX-2 leads to inflammation and act as the prime target (in addition to COX-1) for inhibitions. Selective COX-2 inhibitors such as celecoxib and rofecoxib have proven effective in treating acute pain, osteoarthritis, and rheumatoid arthritis, and constitute severe cardiovascular effects revealed from clinical tests (Vane & Botting, 1998). Drug discovery efforts based on the diaryl heterocycle scaffold with sulfonamide moiety (e.g., Dup-697; chemical name: 5-bromo-2-(4-fluorophenyl)-3-(4-(methylsulfonyl)phenyl)-thiophene) have been studied extensively (Bosch et al., 2000; Kim, Chae, Yi, Park, & Yoo, 2004). This COX-2 dataset was chosen to explain the objective of this study in detail. In addition, the robustness of this approach was evaluated using ten selected diverse datasets including apoptosis inducing 4-aryl-4H-chromenes (Sciabola, Carosati, Cucurull-Sanchez, Baroni, & Mannhold, 2007), azaheterocycles derivatives with anti-malarial activity (Sharma, Sharma, Sharma, & Kumar, 2013), azole antifungal dataset (Talele, Kulkarni, & Kulkarni, 1999), cyclin-dependent kinase 2 (CDK2) inhibitors (Babu, Smiles, Narasu, & Srinivas, 2008), cytochrome P450 2B6 (CYP2B6) inhibitors (Roy & Roy, 2010a), aromatase or cytochrome 19 (CYP19) inhibitors (Roy & Roy, 2010b), HIV-1 reverse transcriptase (HIV-1 RT) inhibitors (Roy & Leonard, 2004), UDP-3-O-[R-3-hydroxymyristoyl]-GlcNAc deacetylase (LpxC) inhibitors (Kadam & Roy, 2006), Plasmodium falciparum protein kinase 7 (PPIK7) inhibitors (Sahu, Shahi, Sharma, & Kohli, 2011) and phosphodiesterase IV inhibitors (Chakraborti, Gopalakrishnan, Sobbia, & Malde, 2003).

Literature survey indicates that majority of the field-based QSAR models were largely affected by the variable selection methods and molecular treatment such as the force field parameters for steric field descriptors and atomic partial charge calculation methods for electrostatic fields, besides the primary limitations of appropriate ‘bioactive’ conformational sampling and alignment (Kubinyi, 1997a, 1997b). The choice of suitable and accurate electrostatic potentials assignment scheme is a crucial task in QSAR work as there are no strict guidelines for the selection of charge models (Tsai et al., 2010). The most commonly used molecular mechanics (MM) force fields (empirical and semiempirical) to calculate MIFs facilitates the development of robust QSAR models. However, the high-throughput assignment of atomic partial charge to molecules via quantum mechanics (QM) approach (ab initio methods) are time consuming and computationally demanding (Cruciani, 2006; Tsai et al., 2010). Most of the CoMFA- and CoMSIA-based QSAR studies that focused on reporting the most
appropriate charge calculation schemes based on statistical measures (e.g., \( r^2 \), \( q^2 \)) do not provide consensus guidelines rather than the recommendation of superior quality of Q SAR models derived from QM and/or semiempirical approaches in comparison with MM methods (Choo, Jung, Koh, & Pae, 2003; Kroemer, Hecht, & Liedl, 1996; Madhavan, Gadhe, Kothandan, Lee, & Cho, 2012; Tsai et al., 2010). Contrastingly, we were interested in the study of activity-correlating molecular regions by various electrostatic schemes in an understanding that the selection of suitable atomic charge models facilitates the mapping of crucial molecular regions that may be missed or remain unexplored if we rely on single charge model. The latest progress in computational strategies for medicinal plant-based drug discovery, structural bioinformatics, and related methods has proved extremely useful (Kumar, George, Jasrai, & Pandya, 2014; Kumar, Jasrai, Mehta, & Pandya, 2014; Kumar, Pandya, Desai, & Jasrai, 2014). There are quite many charge models, and methods have been used for assigning charges include AM1, AM1-BCC, CFF, Del Re, Gasteiger, Gasteiger–Hückel, Hückel, Merck Molecular Force Field (MMFF), PRODRG, Pullman, and VC2003. Each of them has its own advantages and disadvantages. Knowing the maximum diversity in the target molecules and method of assigning charges, we have selected most commonly used six of them namely, Gasteiger and Marsili (1980), Del Re, Pullman, and Yonezawa (1963), and MMFF (1996), Hückel (1931), Gasteiger and Saller (1985), and Berthod, Giessen-Prettre, and Pullman (1967) available in VLife MDS (2008) and SYBYL suites (2010), to map activity-correlating molecular regions and to obtain a consensus impact of descriptors from charge scheme-based QSAR models, were the primary objectives of this study. The resultant generalized QSAR model was then compared with the Open3DQSAR (2011) model to judge the predictions.

2. Materials and methods

2.1. Dataset

We selected eleven diverse datasets of molecules with biological activities spanning various chemical classes or functionalities with at least 4–5 orders of activity magnitude. These include COX-2 selective inhibitors (Chopra, Gupta, Gupta, & Saluja, 2008), apoptosis inducing 4-aryl-4H-chromenes (Sciabola et al., 2007), azaaunones derivatives with anti-malarial activity (Sharma et al., 2013),azole antifungal dataset (Talele et al., 1999), CDK2 inhibitors (Babu et al., 2008), CYP2B6 inhibitors (Roy & Roy, 2010a), CYP19 inhibitors (Roy & Roy, 2010b), HIV-1 RT inhibitors (Roy & Leonard, 2004), LpxC inhibitors (Kadam & Roy, 2006), PiPK7 inhibitors (Sahu et al., 2011) and phosphodiesterase IV inhibitors (Chakraborti et al., 2003). For example, the COX-2 inhibitor dataset covered 7 chemical classes vide 1,2 diarylmadazole, 1,5 diarylpyrrozole, acyclic 2-alkyl-1,2 diaryl (E) olefins, aryl substituted methyleneaminomethyl (MAOM), derivatives of diarylcyclopentenyl COX-2 inhibitors, 4′-(4-cycloalkyl/aryl-oxazol-5yl)benzene sulfonamides, and heteroaromatic analogues of (2-aryl-1-cyclopentenyl-1-alkylidene)-(aryl)methoxy) amine, as used by Chopra et al. (2008) for pharmacophore modeling (Chopra et al., 2008). The structures of the selected COX-2 inhibitors were given in Supplementary Figure 1, and its biological activity was reported in half maximal inhibitory concentration (IC50) in nM units. The COX-2 inhibitor dataset was classified as highly active (<50 nM), moderately active (50–2500 nM), and inactive (>2500 nM) (Chopra et al., 2008). The IC50 values were then converted into its negative logarithmic units, pIC50 to use as dependent variable in QSAR modeling. This procedure was repeated for every selected dataset of this study, and the representative structures of the selected datasets were given in Supplementary Figure 2.

2.2. Dataset partitioning and alignment

The selection of training and test sets along with the compound number scheme was adopted from respective studies. The training set constituted diverse molecules spanning various chemical classes and encompassed an activity range of 4–5 orders of magnitude. For example, the COX-2 dataset was divided into 24 training and 22 test sets based on Chopra et al. (2008) COX-2 pharmacophore study (Chopra et al., 2008). The structures of the molecules were drawn and 3D optimized using Marvin Sketch 5.11.4 program (ChemAxon LLC) (Marvin (version 5.11.4), 2012) and exported in structure data format (SDF). YASARA Structure (academic license) software (Krieger, Darden, Nabuurs, Finkelstein, & Vriend, 2004) was employed to align molecules using atom-to-atom matching technique as the dataset constituted diverse chemical classes of molecules and do not contain a common maximal common substructure to act as template to guide template-based alignment. Prior to alignment, conformational search was executed using conjugate gradient energy minimization technique (atom-typed using MMFF force field, 10,000 iterations, root mean square gradient = 0.001 kcal/mol) to obtain feasible conformers of molecules as input for optimal molecular alignment and returned as alignment file in single SDF format. This alignment file was procured as input to calculate molecular field descriptors in VLife Molecular Design Suite (MDS) version 4 (VLife Sciences Technologies Pvt. Ltd., India; academic license, 2008), SYBYL Suite version 1.2 (Tripos International, St. Louis; academic license, 2010), and Open3DQSAR (http://www.open3dqsar.org, academic freeware, 2011) chemometric softwares, respectively.
Figure 1. Scatter plot of observed and predicted COX-2 inhibitory activities of the selected QSAR models.

Figure 2. The molecular field plots for various atomic partial charge-based QSAR models on COX-2 inhibitors developed from VLife MDS: (A) Gasteiger–Marsili, (B) Del Re, and (C) MMFF.
2.3. Calculation of molecular field descriptors and commonalities in VLife MDS, SYBYL, and Open3DQSAR softwares

The procedure for developing molecular field-based QSAR follows the standard prototype of CoMFA method. The list of recommendations to enable reproducibility of CoMFA results, given by Hugo Kubinyi was strictly followed (Kubinyi, 2008). We used VLife MDS (2008), SYBYL (2010), and Open3DQSAR (2011) softwares to build field-based QSAR and constituted the following commonalities (otherwise mentioned in the next section) in its respective calculations. A sp³ hybridized carbon probe of charge +1 was used to calculate the interaction energies between probe and ligand atoms in a grid box with the customized grid settings in x (For example, COX-2 dataset: start: −11.0000; end: 17.0000), y (start: −23.0000; end: 19.0000), and z (start: −13.0000; end:15.0000) dimensions using a grid spacing of 2 Å. The distance-dependent dielectric function was invoked with a dielectric constant of 1.0. The electrostatic and steric fields interaction energies were refined by the specification of cutoffs for electrostatic field at 10.0 kcal/mol and steric field at 30.0 kcal/mol, respectively, for inclusion as field descriptors. The steric field was computed between n atoms of ligand and sp³ carbon probe; based on Lennard-Jones’ 6–12 potential using AMBER FF99 (Wang, Cieplak, & Kollman, 2000) van der Waals parameters (Open3DQSAR) and Tripos force field (2006) (VLife MDS and SYBYL CoMFA), respectively.

The Lennard-Jones potential can be expressed in the following equation (Cramer et al., 1988; Wade, 1993),

\[ E_{vdW} = \sum_{i,j=1}^{n} \left( A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} \right) \]

Here, \( r_{ij} \) is the distance between atom i of the molecule and the grid point j where the probe atom is located. \( A_{ij} \) and \( B_{ij} \) are the constants associated with the van der Waals radii of the corresponding atoms.

Additionally, the electrostatic field terms are calculated based on Coulomb potential function and can be generalized (Cramer et al., 1988; Wade, 1993) as follows;

\[ E_{ele} = \sum_{i,j=1}^{n} \frac{q_i q_j}{D r_{ij}} \]

Here, \( r_{ij} \) is the distance between atom i of the molecule and the grid point j where the probe atom is located. \( D \) is the dielectric constant, \( q_i \) is the atomic partial charge of the ligand atom i, and \( q_j \) is the charge of the probe j.

Besides the commonalities in the field treatment of molecules in these softwares, PLS modeling was ensured to develop QSAR models.

2.4. Specifications in VLife MDS, SYBYL, and Open3DQSAR softwares

We described here the complete details of selection of atomic partial charge schemes, variables sampling, and visualization settings implemented in VLife MDS (2008), SYBYL (2010), and Open3DQSAR (2011) softwares. As we compared the various atomic partial charge-based QSAR models developed from VLife MDS and SYBYL, default specifications were given to Open3DQSAR software to construct conventional QSAR model. This step ensured the mapping of variables that explain activity-correlating regions by an external program (here, Open3DQSAR, 2011) and provided explanation of MM-based force fields toward the robustness and predictivity of field-based QSAR models. Thus, Open3DQSAR acted as standalone QSAR tool and rely on the point charge model of Coulomb potentials to gauge electrostatic field descriptors, and also on the Lennard-Jones’ 6–12 potential atom-typed using AMBER FF99 (Wang et al., 2000) van der Waals parameters to compute steric field descriptors.

Various electrostatic potentials including Gasteiger and Marsili (1980), Del Re et al. (1963), and MMFF (1996) were assigned using ‘Charge type’ option in ‘Field Compute’ settings of VLife MDS. After the calculation of field descriptors, the ‘Remove invariant column’ command was given to exclude statistically insignificant descriptors and chosen stepwise variable selection method. This stepwise procedure begins by developing a trial model step by step with a single field variable and to each step, and one field variable is added and subsequently inspect the fit of the model. This step is iterative until no more significant variables remaining outside the model are left. The selection of variables during this process was obliged with the following conditions. The cross-correlation limit between variables was set to 0.5, the number of maximum variables of 10, the term selection criteria of \( q^2 \), variance cutoff of 0.0, selection of autoscaling procedure, the specification of Fischer (F) statistic to evaluate the significance of the selected variables. The highly informative lattice descriptors were represented as spherical balls at the lattice intersection in the VLife MDS workspace, to map the activity-correlating molecular regions and its chemical nature (electropositivity, electronegativity, steric favoring and unfavoring regions) which was studied by its arithmetic sign (+ or −) assigned to descriptors in the final PLS equation.

CoMFA models were built in SYBYL suite (2010) using Hückel (1931), Gasteiger and Saller (1985), and Berthod et al. (1967) charge schemes. The alignment generated from YASARA Structure was procured as a molecular database file in SYBYL suite. CoMFA fields were calculated with ‘Tripos Standard’ field class option.
No smoothing of field variables was applied and dropped electrostatic columns within steric cut-off for each molecule. Using spreadsheet dialog, PLS models were developed with cross-validation ($q^2$) to recognize the optimal number of principal components (PCs) used to develop model. Field variables whose energy variance are less than 2 kcal/mol were omitted using ‘column filtering’ option. Subsequently, PLS models were produced without cross-validation and specified the optimal PCs to obtain $r^2$ value. The CoMFA contour maps generated by accounting the aligned chemical groups in the dataset were used to project over selected molecules (most active and inactive) in the integrated SYBYL molecular viewer.

Subsequent to field descriptor calculations in Open3DQSAR, several variable selection procedures were employed. In the pre-treatment operation phase, zeroing of grid values that was close to 0 (level = 0.05), standard deviation (SD) cutoff that removed variables having a SD less than 2 to improve the signal-to-noise ratio, N-level variable elimination which discarded variables containing different energy values across the different objects to avoid bias in the model, were performed to determine the poorly informative variables and eliminate them. We used block unweighted scaling procedure to normalize multiple sets of field descriptors securing different magnitudes (Kastenholz, Pastor, Cruciani, Haaksma, & Fox, 2000). To develop optimal PLS models, some advanced variable selection methods were also applied which include smart region definition (SRD) (Pastor, Cruciani, & Clementi, 1997) and fractional factorial design (FFD) (Baroni et al., 1993). SRD attempted to group variables based on the mapping of descriptors in 3D space and reduced redundancy that originates from multiple nearby descriptors which encode similar information (critical and collapse distance of 1.0 and 2.0 Å were set). FFD was used to select group of variables (type = leave-many-out; 20 runs) that have largest effect on model predictivity from a pool of descriptors refined by earlier SRD run. The PLS pseudo-coefficients contour map depicting the activity-correlating molecular regions were visualized using PyMOL graphical package (academic license) (PyMOL, 2009). The Open3DQSAR and CoMFA contour maps were produced as scalar products of coefficients and SD related with each column.

2.5. Evaluation of the QSAR models

The PLS method was used to develop QSAR models in its respective software applications. We build PLS models considering the field-based descriptors as independent variables and pIC$_{50}$ as dependent variable. Several statistical measures were studied in PLS analyses for evaluating QSAR model’s quality and predictivity. These include coefficient of determination ($r^2$), leave-one-out (LOO) cross-validation ($q^2$), $F$ test, and number of principal components (PCs). We selected the same set of criteria used by Tsai et al. (2010) to compare different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. A $q^2$ value of at least 0.5 was chosen as the criterion for QSAR robustness. Further, the internal prediction accuracy was assessed by $r^2 \geq 0.5$ and $q^2 \geq 0.5$, for training and test sets, respectively (Tsai et al., 2010).

3. Results and discussion

3.1. Assessment of best scoring QSAR models

The statistical measures $r^2$ and $q^2$ have been conventionally used to assess the best scoring QSAR models (Verma et al., 2010; Wermuth, 1996) wherein PLS method is used to develop relationship between biological endpoints (here, pIC$_{50}$ of selected dataset) and the molecular field descriptors chosen by various variable selection methods of respective QSAR programs (Table 1 and Supplementary Table 1). The $r^2$ value (variance explained by the PLS model) of QSAR models based on Gasteiger–Marsili, Del Re, MMFF, Hückel, Gasteiger–Hückel, and Pullman charge schemes indicated that QSAR models based on Hückel and Gasteiger–Hückel charge schemes outperformed in most of the cases when compared against respective charge-based QSAR models. The internal predictivity and robustness of the QSAR models were examined by $q^2$ parameter, and the best scoring PLS models were selected by the highest $q^2$ value. No consensus was obtained to show the superiority of charge models. Further, the difference between $r^2$ and $q^2$ values of various QSAR models was at least ~4.0. Noticeably, the difference was only 2.90 for Hückel-based charge models. This suggests that charge scheme may be directly related to the molecular dataset and thereby requires examination of all available charge schemes. Although the best scoring QSAR models were selected based on the compliance of guideline related to prediction accuracy (i.e., QSAR model scoring $r^2 \geq 0.5$ and $q^2 \geq 0.5$ (Tsai et al., 2010), all the presented PLS models were statistically significant describing the importance of descriptors in relation to its observed pIC$_{50}$.

The objective of exploring various charge models to recognize activity-correlating molecular regions is comprehensively detailed in the COX-2 inhibitors dataset. The observed and predicted activities of the selected COX-2 inhibitor dataset are listed in Table 2 and graphically plotted in Figure 1. The robustness of this method was studied using ten diverse datasets and briefly explained in the later sections. The contribution of various field-based descriptors studied from VLife MDS and SYBYL suites to build PLS relationship in the selected charge-based QSAR models is given in Supplementary.
**Table 1. Statistics of the best scoring PLS models.**

| Targets                                      | QSAR model    | $r^2$   | $q^2$   | Optimal PCs |
|----------------------------------------------|---------------|---------|---------|-------------|
| **COX-2 inhibitors**                        | Open3DQSAR    | 0.9959  | 0.5285  | 5           |
| Del Re                                       | 0.9432        | 0.9113  | 4       |             |
| Gasteiger–Marsili                            | 0.9744        | 0.9550  | 4       |             |
| MMFF                                         | 0.9326        | 0.9096  | 1       |             |
| Hückel                                       | 0.9249        | 0.8524  | 5       |             |
| Gasteiger–Hückel                             | 0.9428        | 0.8614  | 5       |             |
| Pullman                                      | 0.9252        | 0.9075  | 5       |             |
| **Apoptosis inducing 4-aryl-4H-chromenes**   | Open3DQSAR    | 0.6681  | 0.6092  | 5           |
| Del Re                                       | 0.5587        | 0.5456  | 2       |             |
| Gasteiger–Marsili                            | 0.5956        | 0.5103  | 2       |             |
| MMFF                                         | 0.5531        | 0.5492  | 2       |             |
| Hückel                                       | 0.5911        | 0.5730  | 4       |             |
| Gasteiger–Hückel                             | 0.6077        | 0.5427  | 5       |             |
| Pullman                                      | 0.6046        | 0.5397  | 5       |             |
| **Azaaurones derivatives with anti-malarial**| Open3DQSAR    | 0.9464  | 0.5924  | 5           |
| activity (Dataset 2)                         | Del Re        | 0.7295  | 0.5812  | 2           |
| Gasteiger–Marsili                            | 0.8451        | 0.5700  | 3       |             |
| MMFF                                         | 0.8215        | 0.5530  | 2       |             |
| Hückel                                       | 0.9187        | 0.9080  | 4       |             |
| Gasteiger–Hückel                             | 0.9045        | 0.8996  | 4       |             |
| Pullman                                      | 0.9115        | 0.8822  | 4       |             |
| **Azole antifungal agents** (Dataset 3)      | Open3DQSAR    | 0.9835  | 0.7104  | 5           |
| Del Re                                       | 0.9456        | 0.7249  | 4       |             |
| Gasteiger–Marsili                            | 0.9229        | 0.5110  | 2       |             |
| MMFF                                         | 0.9039        | 0.7722  | 3       |             |
| Hückel                                       | 0.9781        | 0.9756  | 3       |             |
| Gasteiger–Hückel                             | 0.9708        | 0.9486  | 4       |             |
| Pullman                                      | 0.9715        | 0.9688  | 4       |             |
| **Cyclin-dependent protein kinase 2 (CDK2)**  | Open3DQSAR    | 0.9364  | 0.6648  | 5           |
| inhibitors (Dataset 4)                       | Del Re        | 0.7862  | 0.5334  | 3           |
| Gasteiger–Marsili                            | 0.6573        | 0.5775  | 1       |             |
| MMFF                                         | 0.5197        | 0.5017  | 4       |             |
| Hückel                                       | 0.8226        | 0.7771  | 1       |             |
| Gasteiger–Hückel                             | 0.8367        | 0.8358  | 2       |             |
| Pullman                                      | 0.9000        | 0.8820  | 1       |             |
| **Cytochrome P450 2B6 (CYP2B6)** (Dataset 5) | Open3DQSAR    | 0.5217  | 0.5163  | 5           |
| Del Re                                       | 0.7937        | 0.7921  | 3       |             |
| Gasteiger–Marsili                            | 0.6302        | 0.5828  | 1       |             |
| MMFF                                         | 0.7935        | 0.7504  | 3       |             |
| Hückel                                       | 0.9111        | 0.8840  | 1       |             |
| Gasteiger–Hückel                             | 0.9576        | 0.8974  | 1       |             |
| Pullman                                      | 0.9520        | 0.8963  | 1       |             |
| **Aromatase (CYP19)** (Dataset 6)            | Open3DQSAR    | 0.9057  | 0.7438  | 5           |
| Del Re                                       | 0.5996        | 0.5636  | 4       |             |
| Gasteiger–Marsili                            | 0.6232        | 0.5591  | 4       |             |
| MMFF                                         | 0.6769        | 0.6786  | 5       |             |
| Hückel                                       | 0.8512        | 0.8494  | 5       |             |
| Gasteiger–Hückel                             | 0.8511        | 0.8316  | 5       |             |
| Pullman                                      | 0.8382        | 0.8300  | 5       |             |
| **HIV-1 reverse transcriptase (HIV-1 RT)** - 2-anino-6-arylsulfonylbenezonitriles inhibitors** (Dataset 7) | Open3DQSAR | 0.9485  | 0.7593  | 5           |
| Del Re                                       | 0.8391        | 0.7844  | 3       |             |
| Gasteiger–Marsili                            | 0.7884        | 0.7445  | 3       |             |
| MMFF                                         | 0.8854        | 0.8725  | 7       |             |
| Hückel                                       | 0.9099        | 0.8559  | 5       |             |
| Gasteiger–Hückel                             | 0.8839        | 0.8176  | 2       |             |
| Pullman                                      | 0.8846        | 0.7400  | 2       |             |

(Continued)
Table 1. (Continued).

| Targets                                                                 | QSAR model   | $r^2$  | $q^2$  | Optimal PCs |
|-------------------------------------------------------------------------|--------------|--------|--------|-------------|
| *Pseudomonas aeruginosa* deacetylase LpxC inhibitors (Dataset 8)        | Open3DQSAR   | 0.9458 | 0.5091 | 5           |
|                                                                          | Del Re       | 0.8559 | 0.5541 | 2           |
|                                                                          | Gasteiger–Marsili | 0.8765 | 0.5190 | 3           |
|                                                                          | MMFF         | 0.8686 | 0.5399 | 2           |
|                                                                          | Hückel       | 0.7679 | 0.6933 | 1           |
|                                                                          | Gasteiger–Hückel | 0.8296 | 0.7227 | 4           |
|                                                                          | Pullman      | 0.8388 | 0.7904 | 3           |
| Imidazopyridazine derivatives with *Plasmodium falciparum* protein kinase 7 (PiPK7) inhibitors (Dataset 9) | Open3DQSAR   | 0.9566 | 0.6667 | 5           |
|                                                                          | Del Re       | 0.7662 | 0.5645 | 1           |
|                                                                          | Gasteiger–Marsili | 0.8520 | 0.6251 | 3           |
|                                                                          | MMFF         | 0.7990 | 0.6403 | 2           |
|                                                                          | Hückel       | 0.9627 | 0.6729 | 5           |
|                                                                          | Gasteiger–Hückel | 0.9694 | 0.5599 | 3           |
|                                                                          | Pullman      | 0.9623 | 0.5271 | 4           |
| Thieno[3,2-d]pyrimidines with phosphodiesterase IV inhibitory activity (Dataset 10) | Open3DQSAR   | 0.9566 | 0.8033 | 5           |
|                                                                          | Del Re       | 0.9618 | 0.5004 | 4           |
|                                                                          | Gasteiger–Marsili | 0.9467 | 0.7972 | 4           |
|                                                                          | MMFF         | 0.9507 | 0.6192 | 4           |
|                                                                          | Hückel       | 0.9440 | 0.8726 | 4           |
|                                                                          | Gasteiger–Hückel | 0.9512 | 0.9000 | 3           |
|                                                                          | Pullman      | 0.9492 | 0.9024 | 4           |

Table 1 and Supplementary Figures 3 and 4, respectively. The observed and predicted activities for the ten datasets are listed in Supplementary Table 2 and graphically shown in Supplementary Figure 5, respectively.

3.2. COX-2 inhibitory activity-correlating molecular regions

The COX-2 inhibitory activity-correlating molecular regions based on the selected 3D alignment of dataset were studied by the respective mapped field descriptors from the selected charge-based QSAR models. The molecular regions were annotated based on the chemical structure of the most active COX-2 inhibitor (4-[4-fluoro-5-phenyl-3-(trifluoromethyl)-1H-pyrazol-1-yl]benzene-1-sulfonamide, 1, IC$_{50}$ = 1.7 nM) and were divided into three regions, viz. region around (i) trifluoromethane group, (ii) sulfonamide moiety, and (iii) functional groups connected to pyrazole core structure other than regions (i) and (ii), respectively, to facilitate the mapping of consensus field descriptors (Figures 2 and 3).

3.3. Set of various charge-based MIF descriptors mapped in the region around trifluoromethane group

One pair of consensus electronegative field descriptors was mapped by the Del Re (E$_{1998}$) and MMFF (E$_{1751}$)-based charge models around trifluoromethane group with a unique electronegative (E$_{1986}$ of Gasteiger–Marsili model) and two unique steric (S$_{1726}$ and S$_{2517}$ of Del Re model) descriptors. The consensus descriptors, E$_{1998}$ and E$_{1751}$, signified the importance of electronegative groups in this region and mapped the electropositive groups present in the COX-2 moderately active molecules. E$_{1998}$ showed the electropositive groups connected with heterocycles which should be avoided to develop active COX-2 inhibitors. This descriptor mapped CH$_3$ (7), neutral charge of cyclohexane (8) and fused aromatic groups with S=O (13) and –O–((CH$_2$)$_4$) (24), respectively. Similar to E$_{1998}$ of Del Re model, E$_{1751}$ descriptor of MMFF model projected above the electropositive moieties vide the electropositive cloud above methylthiophene (7), neutral charge of cyclohexane (8), the electropositive region above substituted indole group (11), the neutral system of thiophene (13) connected with fluorine and methoxy groups (adjacent electronegative groups) at 6th and 7th positions, respectively. Conclusively, Del Re and MMFF partial charge schemes pointed out the avoidance of electropositive groups that are present in the moderately active COX-2 inhibitors. The unique electronegative E$_{1986}$ descriptor of Gasteiger–Marsili model showed the electronegativity conferred by CF$_3$ group is beneficial for COX-2 inhibitory activities. This group is present in the most active dataset of COX-2 inhibitors. The contour plots developed from Hückel,
Gasteiger–Hückel, and Pullman charge-based QSAR models revealed an electropositive cloud above the heterocycles in the trifluoromethane region as noted by E_1998 (Del Re) and E_1751 (MMFF) descriptors. However, an extensive electropositive (blue) contour patch was observed in Hückel model compared to Gasteiger–Hückel and Pullman CoMFA models, respectively. Similar to the presence of electronegative descriptor pair (E_1998 and E_1751), the projection of fluorine atoms from trifluoromethane group in the aligned space generated two corresponding electronegative (red) small contours in the entire CoMFA models.

### Table 2. Observed and predicted COX-2 inhibitory activities of the selected QSAR models.

| Compounds | Observed pIC50 | Open3DQSAR Conventional model | Atomic partial charge-based models (VLife MDS) | Atomic partial charge-based models (CoMFA) | Predicted pIC50 |
|-----------|----------------|-------------------------------|-----------------------------------------------|-------------------------------------------|----------------|
| 1         | 8.77           | 8.514                         | 8.522                                        | 8.436                                     | 8.524          |
| 2         | 8.432          | 8.456                         | 8.631                                        | 8.665                                     | 8.273          |
| 3         | 8.222          | 8.233                         | 7.857                                        | 8.337                                     | 8.062          |
| 4         | 8.097          | 8.242                         | 7.971                                        | 7.64                                      | 8.194          |
| 5         | 7.921          | 7.928                         | 7.894                                        | 7.938                                     | 7.635          |
| 6         | 7.699          | 7.707                         | 7.888                                        | 7.478                                     | 7.538          |
| 7         | 7.585          | 7.592                         | 7.468                                        | 8.066                                     | 7.43           |
| 8         | 7.071          | 7.089                         | 7.452                                        | 7.034                                     | 8.028          |
| 9         | 5.721          | 5.651                         | 5.593                                        | 5.602                                     | 5.928          |
| 10        | 6.292          | 6.284                         | 6.352                                        | 6.501                                     | 6.107          |
| 11        | 6.187          | 6.215                         | 6.507                                        | 6.516                                     | 5.353          |
| 12        | 6.114          | 6.140                         | 5.747                                        | 5.353                                     | 6.224          |
| 13        | 5.824          | 5.907                         | 5.878                                        | 5.648                                     | 5.794          |
| 14        | 5.77           | 5.920                         | 5.784                                        | 5.738                                     | 5.056          |
| 15        | 5.585          | 5.592                         | 5.817                                        | 5.828                                     | 5.804          |
| 16        | 5.509          | 5.448                         | 5.465                                        | 5.683                                     | 6.062          |
| 17        | 5.102          | 5.000                         | 4.661                                        | 4.957                                     | 4.886          |
| 18        | 5.097          | 5.055                         | 4.943                                        | 5.326                                     | 5.186          |
| 19        | 4.921          | 4.889                         | 4.815                                        | 4.584                                     | 4.793          |
| 20        | 4.509          | 4.650                         | 4.673                                        | 4.666                                     | 4.53           |
| 21        | 4.367          | 4.270                         | 4.818                                        | 4.24                                      | 4.495          |
| 22        | 4.167          | 4.304                         | 4.301                                        | 4.803                                     | 4.694          |
| 23        | 4.155          | 4.056                         | 4.089                                        | 4.306                                     | 4.452          |
| 24        | 8.432          | 7.600                         | 7.447                                        | 8.84                                      | 8.399          |
| 25        | 8.328          | 7.639                         | 7.95                                         | 8.569                                     | 8.227          |
| 26        | 8.155          | 7.309                         | 8.071                                        | 7.046                                     | 6.83           |
| 27        | 7.886          | 7.649                         | 6.358                                        | 6.842                                     | 7.76           |
| 28        | 7.824          | 7.761                         | 6.919                                        | 6.982                                     | 9.534          |
| 29        | 7.398          | 5.646                         | 7.3                                          | 6.573                                     | 5.414          |
| 30        | 7.387          | 7.849                         | 8.778                                        | 8.831                                     | 8.338          |
| 31        | 7.824          | 7.946                         | 8.16                                         | 8.615                                     | 8.231          |
| 32        | 6.921          | 6.634                         | 5.559                                        | 4.561                                     | 5.615          |
| 33        | 6.866          | 6.462                         | 6.659                                        | 4.463                                     | 5.518          |
| 34        | 5.772          | 6.397                         | 6.834                                        | 4.377                                     | 5.447          |
| 35        | 6.398          | 6.457                         | 5.651                                        | 5.368                                     | 6.771          |
| 36        | 5.77           | 5.900                         | 4.256                                        | 3.344                                     | 6.103          |
| 37        | 5.77           | 6.089                         | 6.986                                        | 5.454                                     | 4.68           |
| 38        | 5.77           | 5.943                         | 7.033                                        | 5.113                                     | 4.94           |
| 39        | 6.432          | 5.943                         | 7.033                                        | 5.113                                     | 4.94           |
| 40        | 6.046          | 6.618                         | 4.846                                        | 5.566                                     | 5.74           |
| 41        | 5              | 7.300                         | 5.674                                        | 2.376                                     | 7.469          |
| 42        | 4.684          | 5.267                         | 6.464                                        | 4.767                                     | 5.304          |
| 43        | 5.32           | 6.622                         | 6.669                                        | 7.423                                     | 6.071          |
| 44        | 3.699          | 5.423                         | 6.64                                         | 5.531                                     | 7.067          |
| 45        | 3.03           | 5.354                         | 5.297                                        | 4.918                                     | 5.464          |

Note: The observed and predicted activities of ten selected datasets are presented in Electronic Supplementary Information.

*Training set: 1–24, test set: 25–46.*

*Non-compliance of field variables with activity when chemical groups aligned in respective positions.*
The steric descriptor $S_{1726}$ depicted the preference of less bulky groups such as CF$_3$ in this region. Indirectly, $S_{2517}$ of Del Re model represented the bulky substituents such as methylthiophene ($7$) and fluorine and methoxy groups (6th and 7th positions) attached with thiochromene ($13$), are the less favored functional groups. Varied contour magnitude (size and shape) for the steric disfavored region (yellow) was graphically illustrated for chemical connectivities around trifluoromethane group with commonality in steric favorness (green) for CF$_3$ group in all the CoMFA models. Collectively, the MIF descriptors mapped in this region around trifluoromethane group demonstrated the preference of less bulky electronegative groups. In addition, the electropositive and neutral heterocycles that are bulky groups observed in the moderately active COX-2 inhibitors were highlighted.

### 3.4. Set of various charge-based MIF descriptors mapped in the region around sulfonamide moiety

The aligned chemical groups around sulfonamide moiety constituted a unique steric and two pairs of consensus electrostatic MIF descriptors. The consensus electropositive descriptor pair was contributed from Gasteiger–Hückel (E$_{2034}$)- and MMFF (E$_{2281}$)-based charge models, whereas the consensus electronegative descriptor pair was derived from Gasteiger–Marsili (E$_{1254}$)- and Del Re (E$_{2306}$)-based charge models. The E$_{2034}$ descriptor mapped to NH$_2$ group (an electropositive) present in the sulfonamide moiety from the most active COX-2 dataset, which enhances activity. This corresponded to H bond donor (HBD) feature of Chopra et al. (2008) pharmacophore hypothesis (Chopra et al., 2008). Similar information was encoded by E$_{2281}$ descriptor of MMFF model. The Gasteiger–Hückel- and Pullman–CoMFA models exhibited an electropositive contour above NH$_2$ group whereas Hückel-charge model did not highlight this requirement. This is one among the important points that classical Hückel charges may not comprehensively account the electronic distribution of the molecular system where extended Hückel semiempirical models can be advocated. This observation reiterates the examination of more empirical charge schemes.

Figure 3. The molecular field plots for various atomic partial charge-based QSAR models on COX-2 inhibitors developed from SYBYL CoMFA: Hückel, Gasteiger–Hückel, and Pullman.
when interpreting SAR data and development of QSAR models over the preference of convenience and speed in calculations. Conclusively, the NH₂ group with electronegative property influences COX-2 inhibitory activity. In pharmacophore terms, HBD is preferred in this region. Further, the E₁₂₅₄ from consensus descriptor pair projected above the S=O (an electronegative) functional group of sulfonamide moiety from most active COX-2 molecules. This chemical group is encoded as HBA feature in the pharmacophore model of Chopra et al. (2008). Similarly, this information was mapped by E₂₃₀₆ descriptor in its Del Re model. Hence, the S=O group possessing an electronegative character increases the COX-2 inhibitory activity. In pharmacophore terms, HBA is preferred in this region. The CoMFA contour maps from Hückel, Gasteiger–Hückel, and Pullman models represented the steric favorness of sulfonamide moiety in COX-2 inhibitors. Steric disfavorness was noticed in molecules that expand beyond the sulfonamide group into the aligned space majorly from the adjacent groups connected to pyrazole scaffold. Noticeably, the pharmacophore model of Chopra et al. (2008) failed to account both of these features (HBA and HBD) in a single model due to the artifact in the interfeature spacing parameter in Catalyst/HypoGen program (Catalyst, 2008; Chopra et al., 2008). Additionally, the unique steric descriptor S₂₅₅₃ of MMFF model described the bulkiness of sulfonamide group as an important sub-structure for enhanced COX-2 inhibitory activities. Moreover, the most active molecules from training and test sets of COX-2 inhibitors especially contained this group. Collectively, the combination of electropositive and electronegative properties of sulfonamide moiety with steric bulkiness influences COX-2 inhibitory activities.

### 3.5. Set of various charge-based MIF descriptors mapped in the region connected to pyrazole core structure excluding regions (i) and (ii)

This region contained a triplet of consensus steric descriptor derived from the various best scoring charge models, along with three unique electrostatic descriptors, respectively. The consensus triplet descriptors were S₁₄₇₂ (Gasteiger–Marsili model), S₉₄₂ (Del Re model), and S₉₄₂ (MMFF model) with negative coefficients. The S₁₄₇₂ descriptor indicated that the chemical complexities due to substitutions at phenyl group (1) attached to pyrazole core structure can decrease activity. Similar to S₁₄₇₂ descriptor of Gasteiger–Marsili model, the Del Re S₉₄₂ descriptor mapped the substitutions at heterocycles connected to pyrazole group, which include benzylxoy (14), benzodioxole (18), toluene (21), and ethoxybenzene (22) groups from moderately active and inactive COX-2 inhibitor dataset. This descriptor information was also encoded by S₉₄₂ of MMFF model. The CoMFA models showed variations in steric requirements of the COX-2 inhibitors. Steric disfavorness over chemical groups that were combinatorially explored in place of the phenyl system (compared to compounds 1 and 2) was exclusively depicted in the Hückel charge-based model. The Gasteiger–Hückel and Pullman indicated steric favorness of the aromatic group with fewer complexities. Conclusively, these consensus descriptors outlined the chemical complexities connected to pyrazole core structure through substitutions, which should be avoided to increase COX-2 inhibitory activity. The unique electronegative E₁₂₀₀ descriptor of Gasteiger–Marsili model showed the preference of electronegative group at the lattice point 1200, which mapped the neutral benzylxoy group connected to thiophene core group (14) that decreases activity. The E₁₂₀₂ of Del Re model demonstrated that the relatively less electronegative groups compared to phenyl moiety of compound 1 are preferred. This descriptor mapped phenyl moiety (1, 2 and 4), difluoro-isothiochromene group fused with pyrazole system (3) and 3-bromo-2,5-dihydrothiophene moiety (5) that conferred electropositive or neutral regions from most active set. Relatively less electronegative substitutions are preferred in the sulfur-containing pyrazole fused heterocycles (3 and 6) as mapped by the E₁₁₉₀ of MMFF model. Similar to earlier CoMFA models discussion, the electronegativity from the most active set is emphasized in Hückel model whereas electropositivity is highlighted in the moderately active set by Gasteiger–Hückel model. Pullman-based charge model did not suggest any role of electropositivity and/or electronegativity via contour map depiction. Collectively, the complex substitutions connected to pyrazole core group must be avoided and alternatively, less electronegative groups can be substituted as connector to pyrazole core structure.

### 3.6. Division of Open3DQSAR contour map to analyze descriptors from various charge models

The contour plot obtained from developed field-based QSAR model using Open3DQSAR program was divided into three regions as above mentioned. The descriptors mapped by various charge scheme-based models were compared with Open3DQSAR contour map to judge the predictions. The activity influencing electropositive and electronegative contours were shown in yellow and green colors, respectively. The steric favoring and unfavoring regions were depicted in red and blue colors (Figure 4).

### 3.7. Interpretation of contour plot in the region around trifluoromethane group

Compounds with bulky substituent should possess electropositive functional groups such as benzene (9)
heptane (12) toward COX-2 activity enhancement. A small yellow patch corresponded to electropositive surface above pyrazole core group. This contour mapped at carbon atoms of pyrazole system excluding the electronegative center of N atoms and these features are common in most of the COX-2 inhibitors. This contour information agrees with the electronegative field descriptors, viz. E_1998 (Del Re), E_1751 (MMFF) and red contours (Hückel, Gasteiger–Hückel, and Pullman), respectively. Further, the position of CF3 on pyrazole core structure is influential in COX-2 activity enhancement as steric positive (red) contour field.

3.8. Interpretation of contour plot in the region around sulfonamide moiety

Similar to E_2034 (Gasteiger–Marsili model) and E_2281 (MMFF model), the electropositive contour mapped the NH2 group of sulfonamide moiety. Further, the electronegativity conferred by S=O was shown in green contour which matched the descriptor information of E_1254 (Gasteiger–Marsili model) and E_2306 (Del Re model). The bulkiness of sulfonamide moiety mapped by S_2553 (MMFF model) is influential in the most active COX-2 inhibitors. The CoMFA models depicted similar steric and electrostatic descriptors information.

3.9. Interpretation of contour plot in the region connected to pyrazole core structure excluding regions (i) and (ii)

The electropositivity conferred by heterocycles is shown in yellow contour, which correlate with the E_1202 descriptor of Del Re model.

3.10. Application of this approach to ten selected datasets

We selected an additional dataset of ten diverse datasets with support of available pharmacophore and QSAR studies to evaluate the robustness of this approach and briefly explained in this section. The VLife MDS, SYBYL CoMFA, and Open3DQSAR coefficient plots are shown in Figures 5–9, respectively.

3.10.1. Apoptosis inducing 4-aryl-4H-chromenes

Sciabola et al. (2007) utilized triplets of pharmacophoric points (TOPP) descriptors to study SAR of 4-aryl-4H-chromenes having apoptosis inducing effect. The combination of pharmacophoric points, viz. dry-dry-HBA, dry-HB donor-HB acceptor and dry-HB acceptor-HB acceptor, is most prevalent in potent molecules and showed the importance of substitutions at 3rd and 5th positions of phenyl and pyridyl rings (Sciabola et al., 2007). The steric favoring descriptors S_944 (Del Re model) described the importance of bulkiness at 7-amine position of chromene scaffold. The electropositivity conferred by chromene scaffold (Gasteiger–Hückel and Pullman models), whereas the electrostatic negative potential of chromene oxygen (Open3DQSAR) is crucial for apoptosis inducing behavior. Additionally, the 3rd position of benzodiozole is an important determinant where bulkiness (S_484 - Del Re and MMFF models) and electronegativity (Open3DQSAR) are dominant. Adjacently, steric affecting descriptor (S_295 – all VLife models) indicated bulkiness beyond methyl group (2nd position of benzodiozole) drastically reduces the activity. Bulky electropositive (S_235 - Del Re and MMFF; CoMFA models) along with intermediary electronegative oxygen (E_156 - Gasteiger–Marsili, CoMFA, and Open3DQSAR) is activity governing molecular regions.

3.10.2. Azaaurones with anti-malarial activity

In strong association with the pharmacophore model proposed by Sharma et al. (2013), the electron donating groups depicted by E_632, a unique descriptor of Gasteiger–Marsili model, plays a crucial role in governing reactivity of the azaaurones series due to its less steric bulkiness (CoMFA models) characteristics (Sharma et al., 2013). The oxygen atom in the aurone scaffold
Figure 5. The molecular field plots for Del Re, Gasteiger–Marsili, and MMFF charge-based QSAR models developed from VLife MDS on selected datasets: (A) Apoptosis inducing 4-aryl-4H-chromenes, (B) Azaaurones derivatives with anti-malarial activity, (C) Azole antifungal dataset, (D) CDK2 inhibitors, and (E) CYP2B6 inhibitors.

Figure 6. The molecular field plots for Del Re, Gasteiger–Marsili, and MMFF charge-based QSAR models developed from VLife MDS on selected datasets: (A) CYP19 inhibitors, (B) HIV-1 reverse transcriptase (HIV-1 RT) inhibitors, (C) LpxC inhibitors, (D) PiPK7 inhibitors, and (E) Phosphodiesterase IV inhibitors.
Figure 7. The molecular field plots for Hückel, Gasteiger–Hückel, and Pullman charge-based QSAR models developed from SYBYL CoMFA on selected datasets: (A) Apoptosis inducing 4-aryl-4H-chromenes, (B) Azaaurones derivatives with anti-malarial activity, (C) Azole antifungal dataset, (D) CDK2 inhibitors, and (E) CYP2B6 inhibitors.

Figure 8. The molecular field plots for Hückel, Gasteiger–Hückel, and Pullman charge-based QSAR models developed from SYBYL CoMFA on selected datasets: (A) CYP19 inhibitors, (B) HIV-1 reverse transcriptase (HIV-1 RT) inhibitors, (C) LpxC inhibitors, (D) PIPK7 inhibitors, and (E) Phosphodiesterase IV inhibitors.
was mapped by entire CoMFA models. Further, the steric favororness by small lipophilic electron-withdrawing groups was demonstrated by CoMFA and identical electrostatic descriptors of VLife MDS models (E_1099 from Del Re and E_1100 from Gasteiger–Marsili and MMFF models). The electronegative bulkiness (E_979 from Del Re model, entire CoMFA; Open3DQSAR) at 3rd position of benzene connected azaaurone scaffold where functional groups extending beyond propyl and cyclic groups decreases activity. The less steric region at 5th position of benzene should be exploited with diverse chemical groups as designated by identical descriptor S_664 of entire VLife MDS and Gasteiger–Hückel charge models. Additionally, less electronegative substitutions as developed by H and CH₃ groups (E_524 from MMFF and entire CoMFA models) are convenient groups to obtain increased anti-malarial potency.

3.10.3. Azoles with antifungal activity

Talele et al. (1999) developed CoMFA models on azole dataset using Austin model 1 (AM1) charge scheme (Talele et al., 1999). The activity trend in the chosen dataset indicates a linearly connected heterocyclic (up to 4) groups in the molecules correlate with antifungal activity where the terminal groups such as imidazole and piperazine contribute to bulkiness (S_2308 from Del Re; S_1411 from entire VLife MDS models) with electrostatic polarizations, viz. E_1775 (Del Re model) and Open3DQSAR model marking N atoms and aromatic system. Such long lipophilic moieties represent correlation with major hydrophobic pocket of cytochrome P-450_{1a2DM} (Talele, Hariprasad, & Kulkami, 1998). The most important insight is the observation of identical steric disfavoring descriptor from all QSAR models projected above the phenyl, thiethyl, benzo thromephene, and

Figure 9. PLS pseudo-coefficients contour plot of Open3DQSAR model of ten datasets visualized using PyMOL graphical package. (A) Apoptosis inducing 4-aryl-4H-chromenes, (B) Azaaurones derivatives with anti-malarial activity, (C) Azole antifungal dataset, (D) CDK2 inhibitors, (E) CYP2B6 inhibitors, (F) CYP19 inhibitors, (G) HIV-1 reverse transcriptase (HIV-1 RT) inhibitors, (H) LpxC inhibitors, (I) PiPK7 inhibitors, and (J) Phosphodiesterase IV inhibitors.
naphthyl rings present in the least active molecules. Similar to mapping of azole nucleus by electronegative patches (Open3DQSAR), the unique electrostatic descriptor E_1436 from Gasteiger–Marsili model points out this requirement.

### 3.10.4. CDK2 inhibition based on pyridines and pyrimidines scaffold

QSAR study using diverse series of pyridines and pyrimidines scaffolds for CDK2 inhibition showed that electron-withdrawing properties on this scaffold with HBD and steric attributes may help to design potent CDK2 inhibitors with special emphasis on the reduction of HBAs and degree of branching emanating from molecule’s center (Babu et al., 2008). Relatively less electronegative substitutions are preferred in relation to oxygen and sulfur containing groups above pyridine and pyrimidine core structures enhance CDK2 activity which represents bulkiness (S_2716 from MMFF model) with electronegative characteristics (CoMFA models). Descriptors that map scaffold represent the electronegative properties of its N atom (E 827 from Del Re; E 989 from MMFF models) where the hydrophobic substituents (S_140 from Del Re; CoMFA and Open3DQSAR models) may hinder CDK2 potency. The Open3DQSAR model in this case fails to demonstrate prominent regions possibly due to field variable cutoff.

### 3.10.5. CYP2B6 inhibitors set

The pharmacophore hypothesis developed by Roy and Roy (2010a) showed that an electron-rich pyridine group and its connected aromatic groups such as phenyl and naphthalene that are encoded as the HBA and ring features to study CYP2B6 inhibition (Roy & Roy, 2010a). Electropositive descriptors such as E_786 from Del Re, E_809 from Gasteiger–Marsili, E_883 from MMFF models, and CoMFA contour region projected above N containing heterocycles dictates relative less electronegative groups are preferred, whereas its terminal must be explored further by bulky functional groups as indicated by S_1015 of Del Re, CoMFA, and Open3DQSAR models, respectively. This steric requirement holds similar notion on both of the terminal regions with variable substitution sites are preferable with at least two heterocycles connected core groups.

### 3.10.6. CYP19 inhibitors

Roy and Roy (2010b) performed docking and 3D-QSAR studies on a diverse dataset of aromatase (CYP19) inhibitors and concluded the role of one or two HBAs with optimal hydrophobicity as a primary activity governing properties (Roy & Roy, 2010b). The bulkiness around imidazole scaffold is negatively influencing CYP19 activity as mapped by identical pairs of steric descriptors, S_1852 (Del Re and Gasteiger–Marsili models) and Open3DQSAR coefficient maps. Further, the electron-withdrawing properties of the substitution sites are favorable as demonstrated by E_826 and E_1096 (Gasteiger–Marsili model) and E_780 (MMFF model), respectively. Similar details were also mapped by Open3DQSAR model. Scaffold such as imidazole and indole groups largely contributes to electronegative potential. The preference of NO₂ and CN are also represented by electronegative descriptor clusters (electrostatic descriptor placed nearby in VLIFE MDS models), whereas the Open3DQSAR model provided hints on more speculative regions with small polyhedra contour region.

### 3.10.7. Anti-HIV-1 RT activity of 2-amino-6-aryl sulfonylbenzonitriles and its related thio- and sulfinyl congeners

Roy and Leonard (2004) studied SAR of 2-amino-6-aryl sulfonylbenzonitriles and its thio- and sulfinyl congeners having anti-HIV-1 reverse transcriptase (HIV-1 RT) activity tested on MT-4 cell lines (Roy & Leonard, 2004). Similar to trifluoromethane’s detrimental property as noted in the SAR study, the electronegativity (E_643 from both Del Re and Gasteiger–Marsili; CoMFA and Open3DQSAR models) conferred along with its bulkiness (S_525 from Del Re; CoMFA and Open3DQSAR model) reduces activity. The steric (S_532 from entire VLIFE MDS models) and electronegative (E_643 from Del Re and Gasteiger–Marsili models; entire CoMFA models) properties were shown by the sulfone moiety in the scaffold which possesses HIV-1 RT activity enhancing characteristics. A unique descriptor S_176 from Del Re model showed the prominent role of ortho-methoxy in HIV-1 RT activity. These Open3DQSAR models revealed the effect of substitutions at ortho position and presence of sulfone moiety with distribution of CF₃ group can increase HIV-1 RT potency.

### 3.10.8. LpxC inhibitors

The cell wall biosynthetic enzyme LpxC is an attractive target for Pseudomonas aeruginosa therapeutic applications (Kadam & Roy, 2006). A unique steric incremental descriptor S_1190 from all VLIFE and CoMFA models showed fluorophenyl group with electronegative properties based on associated halogen atoms (e.g., Cl) plays important core functionality in LpxC inhibitors. Additionally, hydrophobic carbon chains attached to this core group indicates steric disfavorness (Open3DQSAR). Further, another unique steric descriptor, S_1347 (Del Re and Gasteiger–Marsili), S_1344 (MMFF), and CoMFA contour patches represents extensive branching of O and
N atoms drastically reduces activity where halogenation at either C2 or C3 positions enhances LpxC potency. These observations may be related to the utmost requirement of less electron density at ligand-receptor site (Kadam & Roy, 2006).

3.10.9. PfPK7-imidazopyridazine-based inhibitors

Sahu et al. (2011) performed P. falciparum protein kinase 7 (PfPK7) QSAR studies using VLife MDS software (Sahu et al., 2011). Similar to the electropositivity incremental role at 7th position of imidazopyridazine scaffold, the 4-cyanophenyl group correlates with most active molecules of the dataset as shown by E_609 (Del Re), E_572 (Gasteiger–Marsili), E_572 (MMFF) and electropositive contours of CoMFA models. The steric bulkiness produced by fluoropyridine and morpholine connected with phenyl group correlates with decreased PfPK7 inhibitory activity as noted by unique descriptor S_849 (MMFF model), CoMFA and Open3DQSAR models. Electropositivity conferred by aromatic carbons (exclusive of N atoms) was also highlighted in CoMFA models. Additionally, steric groups other than CN such as carboxamide, cyanomethide, hydroxyl, and imidazole deteriorate PfPK7 activity.

3.10.10. Thiienol [3,2-d] pyrimidines with phosphodiesterase IV inhibitory activity

We noticed that the QSAR observation follows the report of CoMFA model developed by Chakraborti et al. (2003). The nitrogen containing pyrimidines that confer electronegative property was mapped by E_844 (Del Re), E_720 (Gasteiger–Marsili and MMFF models) and Open3DQSAR models. Further, the electropositive character of aromatic rings (E_876 from Del Re, CoMFA and Open3DQSAR models) such as 4-phenoxy butyl is detrimental to the activity. In addition, the steric hinder provided by such groups was predicted by S_862 (Gasteiger–Marsili model) and Open3DQSAR also suggests the selection of appropriate electronegative chemical groups with less bulkiness. Further, the thiienol scaffold should be explored further as S_456 (Del Re), S_314 (Gasteiger–Marsili and MMFF model) and Open3DQSAR coefficient patch projected above this group.

3.11. Limitations and future directions

Compared to empirical charge schemes, quantum mechanical (QM) and quantum chemical (QC) descriptors offer a more precise and comprehensive description of electronic effects of the molecules derived from wave function and can be directly applied to study QSAR (Franke, 1984). Even though the calculations are hardware intensive and time consuming, these methods have gained wide acceptance in correlating biological response (Karelson, Lobanov, & Katritzky, 1996). The most popular semiempirical methods are extended Hückel theory (Hoffmann, 1963), complete neglect of differential overlap (CNDO) (Pople, Santry, & Segal, 1965), intermediate neglect of differential overlap (INDO) (Pople, Beveridge, & Dobosh, 1967), modified INDO (MINDO) (Bingham, Dewar, & Lo, 1975), modified neglect of diatomic overlap (MNDO) (Dewar & Thiel, 1977), Austin model 1 (AM1) (Dewar, Zoebisch, Healy, & Stewart, 1985), and parametric model 3 (PM3) (Stewart, 1989). The major strength of quantum-based methods is its accountability to directly characterize the effect of chemical fragments or groups based on the structure and aids in the study of possible mechanisms of action in terms of chemical reactivity. There are inherent errors associated with the mathematical rigourness and its inability to address bulk effects (Karelson et al., 1996).

Applications of quantum-based charge schemes only revealed the superior performance of QSAR models by yielding high q^2 value. For example, Merz-Singh-Kollman (MK)-electrostatic potential (ESP)-derived charges on moracin analogs (Madhavan et al., 2012), consistent force field (CFF) (Tsai et al., 2010) on ten selected datasets, Austin Model 1-Bond Charge Correction (AM1-BCC) on adamantyl derivatives of P-glycoprotein inhibitors (Gadhe et al., 2011), Mulliken charges on paullones class of glycogen synthase kinase 3β inhibitors (Osolodkin et al., 2010), and electrostatic potential-derived charges on benzodiazepine derivatives (Kroemer et al., 1996). Therefore, it seems that the selection of appropriate charge models in QSAR studies is simply based on the selected molecular dataset and currently cannot be prioritized one charge model over another. On the other hand, MM-based methods provide fast descriptors calculations and model development, but may not cover comprehensively the chemical space (Gadhe et al., 2011).

There is a tendency to select QSAR models based on the highest r^2 and q^2 parameters rather than the comparison of QSAR models obtained from various atomic charges and/or force fields pretreatments. A consensus analysis of various descriptors mapped by the top scoring QSAR models will facilitate the crucial understanding of molecular properties necessary to develop potent molecules. It is well known that the assignment of partial charge schemes directly affects the CoMFA QSAR prediction accuracy (Mittal, Harris, McKinnon, & Sorich, 2009). Thus, the generalized QSAR model derived from various charge models will map essential structural properties that may be missed or remain unexplored when relied on single QSAR model.

The current scenario of domination by commercial closed-source programs in the field of 3D-QSAR modeling has been gradually replaced by the open-source packages (Tosco & Balle, 2011). QSAR models
developed from VLife MDS provide a single PLS model with information on optimal PCs used to derive model. However, the details regarding the prediction accuracy (\(r^2\) or \(q^2\)) during the trend in the selection of optimal PCs (a facility provided by Open3DQSAR PLS engine) will be helpful to recognize over-fitting of data points. Further, the manual specification of PCs in the cross-validation PLS model should be automated in SYBYL CoMFA technique. It is an advantage to represent the field-based descriptors as highly informative grid points rather than the contour maps that are directly affected by isosurface field rendering command. A nonlinear modeling equipped with stochastic variable selection methods implemented in commercial QSAR packages drastically reduces the collinearity of variables and over-fitting of the resultant models (Ajmani et al., 2006) (a feature absent in open-source packages). There is a need to implement various advanced grid-based selections in addition to stochastic methods to obtain optimal QSAR models. It should be noted that there are alternative ways (e.g., alignment methods, different probes, chemometric softwares) to develop optimal QSAR models and selection of such models may offer varied interpretations.

4. Conclusions

We studied the consensus activity-correlating molecular regions which were derived from optimal MIF descriptors of the various (best scoring) atomic charge-based QSAR models. The interpretation of MIF descriptors obtained from charge-based QSAR models was carried out by analyzing its highly informative lattice descriptors and subsequently divided into molecular regions. The electrostatic and steric descriptors were gathered region-wise from six atomic charge schemes such as Del Re, Gasteiger–Marsili, MMFF, Hückel, Gasteiger–Hückel, and Pullman, to interpret the requirements of chemical groups in relation to activity. The essential molecular properties mapped by respective MIF descriptors were compared with conventional field-based QSAR model obtained from Open3DQSAR program to validate the predictions. The better performance of Hückel- and Gasteiger–Hückel-based QSAR models was noted in terms of correlation coefficient (\(r^2\)), and the difference between correlation coefficient and its cross-validated terms of \(-0.4\) suggests that the performance of atomic charge schemes to obtain optimal QSAR models may vary across datasets and therefore requires examination of all available MM-based models to arrive at conclusions. The procedure associated with variable selection methods to develop a statistically significant QSAR model precludes the selection of optimal field-based features in a single model. Moreover, the fact that atomic partial charge assignment schemes significantly affects the prediction accuracy of a QSAR model necessitates the study of various charge-based models to retrieve prominent activity governing molecular properties. From this study, it is quite evident that accuracy of charge assignment has a major role to play in developing QSAR models. In such situation, we would like to suggest that one must be careful in selecting the MM-based charge assignment model before taking any conclusion based on the developed QSAR model when quantum mechanical calculations related resources are limited to the user.

Supplementary material
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/07391102.2015.1044474.

Acknowledgments
SPK acknowledges support from Department of Science and Technology, New Delhi, India, as INSPIRE Senior Research Fellowship. PCJ would like to thank University Grants Commission (UGC), Govt. of India for providing starting grant. The authors would like to thank Mr Vijay P. Mehta for help in graphical representations.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by Department of Science and Technology, New Delhi [grant number IF120266], and University Grants Commission (UGC), Govt. of India.

References
Ajmani, S., Jadhav, K., & Kulkarni, S. A. (2006). Three-dimensional QSAR using the \(k\)-nearest neighbor method and its interpretation. *Journal of Chemical Information and Modeling*, 46, 24–31.

Babu, P. A., Smiles, D. J., Narasu, M. L., & Srinivas, K. (2008). Identification of novel CDK2 inhibitors by QSAR and virtual screening procedures. *QSAR & Combinatorial Science*, 27, 1362–1373.

Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., & Clementi, S. (1993). Generating Optimal Linear PLS Estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. *Quantitative Structure-Activity Relationship*, 12, 09–20.

Berthod H., Giessner-Prettre C., & Pullman A. (1967). Sur les rôles respectifs des électrons \(\sigma\) et \(\pi\) dans les propriétés des dérivés halogénés des molécules conjuguées [On the respective roles of \(\sigma\) and \(\pi\) electrons in the properties of halogen derivatives of conjugated molecules]. *Theoretica Chimica Acta*, 8, 212–222. (in French)

Bingham, R. C., Dewar, M. J. S., & Lo, D. H. (1975). Ground states of molecules. XXV. MINDO/3. An improved version of the MINDO semiempirical SCF-MO method. *Journal of the American Chemical Society*, 97, 1285–1293.
Bosch, J., Roca, T., Catena, J. L., Llorens, O., Pérez, J. J., Lagunas, C., ... Farrerons, C. (2000). Synthesis and biological evaluation of 1,3,4-triaryl-3-pyrrolin-2-ones, a new class of selective cyclooxygenase-2 inhibitors. Biorganic & Medicinal Chemistry Letters, 10, 1745–1748.

Catalyst. (2008). San Diego: Accelrys Software. Retrieved from http://accelrys.com/products/discovery-studio/pharmacosphore.html

Chakraborti, A. K., Gopalakrishnan, B., Sobhia, M. E., & Malde, A. (2003). 3D-QSAR studies on antimutary arylsulfonilimidazolidinone derivatives by CoMFA and CoMSIA. Biorganic & Medicinal Chemistry, 11, 4585–4589.

Chopra, M., Gupta, R., Gupta, S., & Saluja, D. (2008). Molecular modeling study on chemically diverse series of cyclooxygenase-2 selective inhibitors: Generation of predictive pharmacophore model using catalyst. Journal of Molecular Modeling, 14, 1087–1099.

Consonni, V., & Todeschini, R. (2010). Molecular descriptors. In T. Puzyn, J. Leszczynski, & M. T. D. Cronin (Eds.), Recent advances in QSAR studies: Methods and applications (pp. 29–102). London: Springer.

Cramer, R. D., Patterson, D. E., & Busch, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of American Chemical Society, 110, 5959–5967.

Crucciani, G. (Ed.). (2006). Molecular interaction fields: Applications in drug discovery and ADME prediction. Weinheim: Wiley-VCH.

Dannhardt, G., & Kiefer, W. (2001). Cyclooxygenase inhibitors – current status and future prospects. European Journal of Medicinal Chemistry, 36, 109–126.

Del Re, G., Pullman, B., & Yonezawa, T. (1963). Electronic structure of the alpha-amino acids of proteins. I. Charge distributions and proton chemical shifts. Biochimica et Biophysica Acta, 73, 153–182.

Dewar, M. J. S., & Thiel, W. (1977). Ground state of molecules, 38. The MNDO method: approximations and parameters. Journal of the American Chemical Society, 99, 4899–4907.

Dewar, M. J. S., Zoebisch, E. G., Healy, E. F., & Stewart, J. J. P. (1985). AM1: A new general purpose quantum mechanical model. Journal of the American Chemical Society, 107, 3902–3909.

Field-Based QSAR (version 2013-3). (2013). New York, NY: Schrödinger, LLC. Retrieved from http://www.schrodinger.com/Field-Based-QSAR/

Franke, R. (Ed.). (1984). Theoretical drug design methods. Amsterdam: Elsevier.

Goodfellow, P. J. (1985). A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. Journal of Medicinal Chemistry, 28, 849–857.

Halgren, T. A. (1996). Merck molecular force field. 2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. Journal of Computational Chemistry, 17, 520–552.

Hoffmann, R. (1963). An extended Hückel theory. I. Hydrocarbons. The Journal of Chemical Physics, 39, 1397–1412.

Hückel E. (1931). Quantentheoretische beiträge zum [Quantum theoretical contributions]. Zeitschrift für Physik, 70, 204–286. (in German)

Kadhim, M., & Roy, N. (2006). Cluster analysis and two-dimensional quantitative structure-activity relationship (2D-QSAR) of Pseudomonas aeruginosa deacetylase LpxC inhibitors. Biorganic & Medicinal Chemistry Letters, 16, 5136–5143.

Karelson, M., Lobanov, V. S., & Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR Studies. Chemical Reviews, 96, 1027–1044.

Kastenholz, M. A., Pastor, M., Crucciani, G., Haaksmma, E. E. J., & Fox, T. (2000). GRID/CPCA: A new computational tool to design selective ligands. Journal of Medicinal Chemistry, 43, 3033–3044.

Kim, H. J., Chae, C. H., Yi, K. Y., Park, K. L., & Yoo, S. E. (2004). Computational studies of COX-2 inhibitors: 3D-QSAR and docking. Bioorganic & Medicinal Chemistry, 12, 1629–1641.

Klebe, G., Abraham, U., & Mietzner, T. (1994). Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. Journal of Medicinal Chemistry, 37, 4130–4146.

Klebe, G., Mietzner, T., & Weber, F. (1994). Different approaches toward an automatic structural alignment of drug molecules: applications to sterol mimics, thrombin and thrombomixins inhibitors. Journal of Computer-Aided Molecular Design, 8, 751–778.

Krieger, E., Darden, T., Nabuurs, S. B., Finkelstein, A., & Vriend, G. (2004). Making optimal use of empirical energy functions: Force-field parameterization in crystal space. Proteins, 57, 678–693.

Kroemer, R. T., Hecht, P., & Liedl, K. R. (1996). Different electrostatic descriptors in comparative molecular field analysis: A comparison of molecular electrostatic and coulomb potentials. Journal of Computational Chemistry, 17, 1296–1308.

Kubinyi, H. (1997a). QSAR and 3D QSAR in drug design, Part 1: Methodology. Drug Discovery Today, 2, 457–467.

Kubinyi, H. (1997b). QSAR and 3D QSAR in drug design, Part 2: Applications and problems. Drug Discovery Today, 2, 538–546.

Kubinyi, H. (2008). Comparative molecular field analysis (CoMFA). In J. Gasteiger (Ed.), Handbook of chemoinformatics: From data to knowledge in 4 volumes (pp. 1555–1574). Weinheim: Wiley-VCH.

Kubinyi, H., Hamprecht, F. A., & Mietzner, T. (1998). Three-dimensional quantitative similarity-activity relationships (3D QSAR) from SEAL similarity matrices. Journal of Medicinal Chemistry, 41, 2553–2564.

Kumar, S. P., Pandya, H. A., Desai, V. H., & Jasrai, Y. T. (2014). Compound prioritization from inverse docking experiment using receptor-centric and ligand-centric resonance concept. Angewandte Chemie, International Edition in English, 24, 687–689.
methods: A case study on Plasmodium falciparum Fab enzymes. *Journal of Molecular Recognition*, 27, 215–229.

Kumar, S. P., Jasrai, Y. T., Mehta, V. P., & Pandya, H. A. (2014). Development of pharmacophore similarity-based quantitative activity hypothesis and its applicability domain: Applied on a diverse data-set of HIV-1 integrase inhibitors. *Journal of Biomolecular Structure & Dynamics*, 33, 706–722.

Kumar, S. P., George, L. B., Jasrai, Y. T., & Pandya, H. A. (2014). Prioritization of active antimalarials using structural interaction profile of Plasmodium falciparum enoyl-acyl carrier protein reductase (PfENR)-triclosan derivatives. *SAR and QSAR in Environmental Research*, 26, 61–77.

Lill, M. A. (2007). Multi-dimensional QSAR in drug discovery. *Drug Discovery Today*, 12, 1013–1017.

Madhavan, T., Gadhe, C. G., Kothandan, G., Lee, K., & Cho, S. J. (2012). Various atomic charge calculation schemes of CoMFA on HIF-1 inhibitors of moracin analogs. *International Journal of Quantum Chemistry*, 112, 995–1005.

Marvin (version 5.11.4). (2012). Hungary: ChemAxon LLC. Retrieved from http://www.chemaxon.com/

Mittal, R. R., Harris, L., McKinnon, R. A., & Sorich, M. J. (2009). Comparative study of E-state parameters. *Chemical Physics*, 47, 653–668.

Miyawaki, T., Tabuse, K., & Sajiki, T. (2011). Development of pharmacophore alignment models as input for comparative molecular field analysis of a diverse set of azole antifungal agents. *Journal of Chemical Information and Computer Sciences*, 39, 958–966.

Pastor, M., Cruciani, G., & Clementi, S. (1997). Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure-activity relationships. *Journal of Medicinal Chemistry*, 40, 1455–1464.

Pople, J. A., Santry, D. P., & Segal, G. A. (1965). Approximate self-consistent molecular orbital theory I. Invariant procedures. *The Journal of Chemical Physics*, 43, S129–S135.

Pople, J. A., Beveridge, D. L., & Dobosh, P. A. (1967). Approximate self-consistent molecular orbital theory V. Intermediate neglect of differential overlap. *The Journal of Chemical Physics*, 47, 2026–2033.

PyMOL. (2009). Palo Alto: DeLano Scientific LLC. Retrieved from http://www.pymol.org.

Roy, K., & Leonard, J. T. (2004). QSAR modeling of HIV-1 reverse transcriptase inhibitor 2-amino-6-arylsulfonylbenzonitriles and congeners using molecular connectivity and E-state parameters. *Bioorganic & Medicinal Chemistry*, 12, 745–754.

Roy, P. P., & Roy, K. (2010a). Pharmacophore mapping, molecular docking and QSAR studies of structurally diverse compounds as CYP2B6 inhibitors. *Molecular Simulation*, 36, 887–905.

Roy, P. P., & Roy, K. (2010b). Docking and 3D-QSAR studies of diverse classes of human aromatase (CYP19) inhibitors. *Journal of Molecular Modeling*, 16, 1597–1616.

Sahu, N. K., Shahi, S., Sharma, M. C., & Kohli, D. V. (2011). QSAR studies on imidazopyridazine derivatives as PIPK7 inhibitors. *Molecular Simulation*, 37, 752–765.

Sciabola, S., Carosati, E., Cucurull-Sanchez, L., Baroni, M., & Mannhold, R. (2007). Novel TOPP descriptors in 3D-QSAR analysis of apoptosis inducing 4-aryl-4H-chromenes: Comparison versus other 2D- and 3D-descriptors. *Bioorganic & Medicinal Chemistry*, 15, 6450–6462.

Stewart, J. J. P. (1989). Optimization of parameters for semi-empirical methods 1-method. *Journal of Computational Chemistry*, 10, 209–220.

SYBYL-X (version 1.2). (2010). St. Louis: Tripos International. Retrieved from http://www.certara.com

Talete, T. T., Hariprasad, V., & Kulkarni, V. M. (1998). Docking analysis of a series of cytochrome P-45014RDM inhibiting azole antifungals. *Drug Design Discovery*, 15, 181–190.

Talete, T. T., Kulkarni, S. S., & Kulkarni, V. M. (1999). Development of pharmacophore alignment models as input for comparative molecular field analysis of a diverse set of azole antifungal agents. *Journal of Chemical Information and Computer Sciences*, 39, 958–966.

Tosco, P., & Balle, T. (2011). Open3DQSAR: A new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *Journal of Molecular Modeling*, 17, 201–208.

Tripos force field, SYBYL. (2006). St. Louis: Tripos International. Retrieved from http://www.certara.com/

Vane, J. R., & Botting, R. M. (1998). Anti-inflammatory drugs and their mechanism of action. *Inflammation Research*, 47, S78–S87.

Vermat, J., Khedkar, V. M., & Coutinho, E. C. (2010). 3D-QSAR in drug design – A review. *Current Topics in Medicinal Chemistry*, 10, 95–115.

Vinter, J. G. (1994). Extended electron distributions applied to the molecular mechanics of some intermolecular interactions. *Journal of Computer-Aided Molecular Design*, 8, 653–668.

VLife MDS (version 4.3): Molecular Design Suite. (2008). Pune: VLife Sciences Technologies. Retrieved from http://www.vlifesciences.com

Wade R. C. (1993). Molecular interaction fields. In H. Kubinyi (Ed.), *3D QSAR in drug design. Theory, methods and applications* (pp. 486–505). Lieden: ESCOM.

Wang, J. M., Cieplak, P., & Kollman, P. A. (2000). How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21, 1049–1074.

Wermuth, C. G. (Ed.). (1996). *The practice of medicinal chemistry*. London: Academic Press.