Development of pharmacophore similarity-based quantitative activity hypothesis and its applicability domain: applied on a diverse data-set of HIV-1 integrase inhibitors

Sivakumar Prasanth Kumar, Yogesh T. Jasrai*, Vijay P. Mehta and Himanshu A. Pandya

Department of Bioinformatics, Applied Botany Centre (ABC), Gujarat University, Ahmedabad 380009, Gujarat, India

Communicated by Ramaswamy H. Sarma

(Received 5 January 2014; accepted 21 March 2014)

Quantitative pharmacophore hypothesis combines the 3D spatial arrangement of pharmacophore features with biological activities of the ligand data-set and predicts the activities of geometrically and/or pharmacophoric similar ligands. Most pharmacophore discovery programs face difficulties in conformational flexibility, molecular alignment, pharmacophore features sampling, and feature selection to score models if the data-set constitutes diverse ligands. Towards this focus, we describe a ligand-based computational procedure to introduce flexibility in aligning the small molecules and generating a pharmacophore hypothesis without geometrical constraints to define pharmacophore space, enriched with chemical features necessary to elucidate common pharmacophore hypotheses (CPHs). Maximal common substructure (MCS)-based alignment method was adopted to guide the alignment of carbon molecules, deciphered the MCS atom connectivity to cluster molecules in bins and subsequently, calculated the pharmacophore similarity matrix with the bin-specific reference molecules. After alignment, the carbon molecules were enriched with original atoms in their respective positions and conventional pharmacophore features were perceived. Distance-based pharmacophoric descriptors were enumerated by computing the interdistance between perceived features and MCS-aligned ‘centroid’ position. The descriptor set and biological activities were used to develop support vector machine models to predict the activities of the external test set. Finally, fitness score was estimated based on pharmacophore similarity with its bin-specific reference molecules to recognize the best and poor alignments and, also with each reference molecule to predict outliers of the quantitative hypothesis model. We applied this procedure to a diverse data-set of 40 HIV-1 integrase inhibitors and discussed its effectiveness with the reported CPH model.

Keywords: Pharmacophore similarity; quantitative pharmacophore hypothesis; pharmacophore space; MCS-based alignment; distance-based pharmacophoric descriptors; HIV-1 integrase inhibitors

Introduction

Three-dimensional pharmacophore models provide an intuitive way of representation for medicinal chemists to understand small molecule binding properties and to interpret chemical-functional molecular characteristics that are necessary to trigger a biological response (Laggner et al., 2005; Steindl, Laggner, & Langer, 2005). It has a tremendous application in drug design by providing valuable information to study structure-activity relationships (SARs) and uncovers the mechanism of ligand-target interactions by deducing the nature of functional groups and non-covalent bonding patterns (Van Drie, 2003; Wermuth & Langer, 1993). It is also applied to the discovery and development of novel molecules with a desired biological activity (Dror, Shulman-Peleg, Nussinov, & Wolfson, 2004). A pharmacophore can be derived on the basis of protein receptor structural information (structure-based approach) or purely on the structures of bioactive and potential ligands (ligand-based approach). Most popular commercial packages utilize ligand-based approach to propose pharmacophore models and also facilitate predictions of biological activities by developing quantitative hypothesis (Patel, Gillet, Bravi, & Leach, 2002). Ligand-based pharmacophore modeling develops hypothesis based on the 3D molecular alignment guided by chemical structure superpositioning, pharmacophore-based overlay, matching shape characteristics, field-based features, etc. (Clement & Mehl, 2000; Cramer, Patterson, & Butane, 1988; Dixon et al., 2006; Guner, 2000; Haigh, Pickup, Grant, & Nicholls, 2005; Li, Sutter, & Hoffmann, 2000).

The official IUPAC definition for ‘pharmacophore’ term states that ‘a pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response’ (Wermuth, Ganellin, Lindberg, & Mitscher, 1998). Pharmacophore pattern or configuration can be defined as a set of relative locations of a pharmacophore feature in 3D space. A pharmacophore
hypothesis is a set of common 3D pharmacophores amongst active compounds in the training set (Dror et al., 2004; Wermuth et al., 1998). This term is also known as common pharmacophore hypotheses (CPHs) models with a modification that each hypothesis is scored according to various geometric and heuristic criteria (Dixon et al., 2006). A finite set of pharmacophores is called pharmacophore space (Dror et al., 2004). Numerous pharmacophore modeling programs extended the meaning of the hypothesis by bridging the geometrical ‘qualitative’ hypothesis described by angles, distances, and planes with experimental biological activities to ‘quantitative’ pharmacophore models with predicted activities (Lemmen & Lengauer, 2000). To address flexibility criteria in diverse data-set, several programs calibrated the hypothesis development phase by incorporating various conditions, including the predefinition of mapping minimum and maximum feature points (Dixon et al., 2006), number of molecules (actives, moderately actives, and less than half of the inactives (Li et al., 2000)) used to develop hypothesis, selecting training molecules with 4 activity orders of magnitude (Li et al., 2000), selecting a set of reference molecules (Schneidman-Duhovny, Dror, Inbar, Nussinov, & Wolfson, 2008), features interdistance flexibilities (VLifeMDS, 2010), etc. The Phase program recommends features-based match for hypothesis generation in case of diverse and/or flexible molecules (Dixon et al., 2006).

Diverse molecules with high chemical complexities may face several problems in pharmacophore modeling for the following reasons, (i) the conformations of the molecules are substantially larger due to the high number of rotatable groups (Kirchmair, Laggner, Wolber, & Langer, 2005), (ii) a simple common substructure in the data-set complicates the alignment or overlaying process (Wolber, Domhofer, & Langer, 2006), (iii) the greater counts of pharmacophore features in each molecule and sampling common features present in entire or partial (cover most and moderately actives) training molecules are exhaustive (Li et al., 2000) and typically constrained by geometric or feature mapping requirements governed by the number of features, tolerance, and distance predefined parameters (Dixon et al., 2006). Thus, mapping pharmacophore space will be very useful for diverse data-sets instead of CPH models to develop quantitative models.

With ongoing efforts on development of computational strategies on medicinal plants-based drug discovery, structural bioinformatics, and related methods (Kumar, Jasrai, Pandya, George, & Patel, 2013; Kumar, Jasrai, Pandya, & Rawal, 2013; Kumar, Pandya, Desai, & Jasrai, 2014), we present here a ligand-based computational procedure for quantitative pharmacophore hypothesis development using pharmacophore similarity matrix calculated on the basis of molecular topology-based alignment technique. The applicability domain of the developed quantitative models was then discussed. Finally, the effectiveness of sampling pharmacophore space was addressed by comparing the quantitative models developed on 40 HIV-1 integrase (HIV-1 IN) inhibitors with its published CPH model (Mustata, Brigob, & Briggsa, 2004). We selected HIV-1 IN inhibitors as HIV-1 IN enzyme is one among the attractive targets in antiviral drug design and expected to complement the therapeutic use of HIV protease and transcriptase inhibitors which have flexible binding affinities (Goldgur et al., 1999; Purohit et al., 2008; Purohit, Rajendran, & Sethumadhavan, 2011; Purohit & Sethumadhavan, 2009).

Materials and methods

Ligand data-set

The ligand data-set consisted of 40 HIV-1 IN inhibitors which covered a wide activity range of .04 μM (compound 11)–1000 μM (compound 15) with structural diversity. A 3D pharmacophore model using Catalyst/ HypeGen program (Catalyst, 2000) based on this diverse data-set was published by Mustata et al. (2004). The biological activity data reported in half maximal inhibitory concentration (IC_{50}) spanning 5 orders of magnitude were converted into its logarithmic scale, pIC_{50} and used as the dependent variable for quantitative pharmacophore modeling. Mustata et al. (2004) selected at least three compounds in each activity order, including the most active and inactive ones, as training set. The data-set was partitioned into training (26 molecules) and test (14 molecules) sets for hypothesis generation.

Maximal common substructure identification

The maximal common substructure (MCS) present in the molecules was identified using ‘Multiple MCS’ program (Tripod, NIH). The program utilizes fragment-based approach that implements fragments as seeds to the C-clique alignment algorithm, to reduce the search space and recognize multiple MCSs present in the molecules (Multiple MCS program, 2010). The entire data-set was specified as input in a single structure data format file and the multiple MCSs were returned.

Molecular topology-based alignment

The molecular data-set was geometrically optimized using AMBER03 force field (Duan et al., 2003) in YASARA Structure software (academic license) (Krieger, Darden, Nabuurs, Finkelstein, & Vriend, 2004). Molecular topology-based alignment was performed for ligand data-set (having ‘carbon’ molecules; described in Results and Discussion) using VLife ‘Align molecules’ module.
acted as a template in its to chosen MCS (present in all training molecules) which acted as a template in its first step, followed by alignment of template-connected chemical groups to obtain optimal alignment (VLifeMDS, 2010). The alignment can be examined graphically and statistically using root-mean-squared deviation (RMSD) parameter, respectively.

**Pharmacophore perception and its descriptors**

The pharmacophore features present in the molecules were predicted using ‘MolSign’ module of VLife MDS v3.5 software (academic license) (VLifeMDS, 2010). MolSign module utilizes a distance geometry method to ‘qualitatively’ predict CPHs. The pharmacophore feature detection procedure was customized to predict the maximum number of features with no tolerance and maximum distance coverage. The process of pharmacophore feature definition, abstraction, and representation in VLife MDS are similar to the existing pharmacophore modeling softwares (VLifeMDS, 2010). The predicted features along with the atomic coordinates for each molecule were saved in XYZ coordinate (.xyz format) file.

A set of distance-based pharmacophoric descriptors were determined by calculating the distance between the feature positions (Cartesian coordinates) and the centroid of the chosen MCS position in the molecule. It should be noted that this descriptor calculation requires molecules in alignment mode. The predicted feature positions saved in XYZ files can be easily recognized by atom labels viz. HAc (H-bond acceptor), HDr (H-bond donor), AroC (aromatic group), AlaC (aliphatic group), PosC (positive ionizable center), and NegC (negative ionizable center). AroC and AlaC are the hydrophobic features.

**Descriptor selection and machine learning**

Since we used flexible parameters to detect pharmacophore features in molecules, the number of features overlaid in the perception phase summed up to 58. Inconsistent pharmacophoric descriptors were identified by inspecting descriptor columns enriched with value zero. The removal of inconsistent descriptors reduced the descriptor set to 9 (HAc – 3, HDr – 2, AroC – 2, AlaC – 2) and used as independent variables in machine learning.

Support vector machines (SVMs) are a famous ‘supervised’ machine learning method for data classification, regression, and other learning tasks (Hsu, Chang, & Lin, 2003). We considered SVM for developing a statistical relationship between the biological activities and pharmacophoric descriptors of the small molecules. The original ligand data-set (40 molecules) was divided into 21 training and 5 validation sets for SVM model training and cross-validation purposes, while the test set used in all these phases was directly employed as an external test set to exhibit SVM model external predictivity. The use of validation set for cross-validation can prevent the model over-fitting problem (Hsu et al., 2003) and selected one compound from each activity in order (i.e. pIC50 ~3, 4, 5, 6, and 7) to ensure activity and structural diversities in the validation set.

LIBSVM (a library of SVMs) package was used to develop quantitative pharmacophore model using computed descriptor set as class attributes (x vectors) and pIC50 values as class labels (target value). The aim of SVM is to generate a model (based on training data) to predict the activities of the test set using test data descriptors by mapping the training x vectors into a higher dimensional space and finding a linear separating hyperplane to achieve marginal and maximum separations between the selected classes (Chang & Lin, 2001; Hsu et al., 2003). Various kernels including linear, polynomial, radial basis function, sigmoid, and anova were selected to obtain an optimal SVM model. A grid search was carried out to monitor various SVM parameters viz. C, mean squared error (MSE), $\epsilon$, $\gamma$, $k_0$, and $r^2$ (coefficient of determination) across models. Based on the recommendations of LIBSVM (Hsu et al., 2003), optimal models were selected on the basis of SVM parameters, cross-validation techniques, and residuals in activities. Statistical analysis of data was performed using SPSS software (SPSS Inc., 2007).

**Results and discussion**

**Pharmacophore similarity-based quantitative pharmacophore hypothesis**

We describe a ligand-based computational procedure for quantitative pharmacophore hypothesis development by proposing geometric flexibilities in pharmacophore model development. Figure 1 gives an outline of the major tasks involved in this procedure and its various stages is discussed in detail in the following sections. The major goal is to map the pharmacophore space necessary to study ligand pharmacophore requirements hidden in highly complex molecules rather than elucidation of CPH models which may differ substantially due to its alignment, pattern search, or heuristic methods (Wolber, Seidel, Bendix, & Langer, 2008). Quantitative pharmacophore hypothesis attempts to develop a statistical relationship of CPH models with biological activities which may cover only a common pharmacophore space shared by aligning chemical groups in the molecular data-set.
Flexible mapping of pharmacophore space is prominent than CPH models for four reasons, (i) molecules with large chemical diversity may bind at different binding sites of a receptor leading to an incorrect pharmacophore model (Dror et al., 2004), (ii) pharmacophore models are mostly dependent upon 3D alignment (overlay) that are carried out by geometric and/or feature matching algorithms (Wolber et al., 2008). Choosing a set of features or neglecting a single or set of feature(s) priori for CPH model generation is questionable in the case of diverse molecules, (iii) CPH models developed on large chemical space do not necessarily indicate complete-ligand binding requirements for a specific receptor (Dror et al., 2004), and (iv) pharmacophore space may encompass many CPH models and mapping its space will be more valuable than selecting top-scored CPH models as it enhances the dimensionality of ligand-binding modes (Cato, 2000; Patel et al., 2002) and contains various chemical groups that are not selected in statistically significant CPH models.

**MCS in the data-set**

We selected 40 HIV-1 IN structurally diverse ligands with its experimental biological activities (pIC_{50}) spanning 5 orders of magnitude. Most of the commercial packages utilize MCS-based alignment to superpose molecules in pharmacophore modeling (Dror et al., 2004; Wolber et al., 2008). It should be noted that the MCS-based and topological-guided alignments (fuzzy search)
are equivalent in performance to clique-detection methods in pharmacophore modeling (Bonachéra & Horvath, 2008; Gardiner, 2013). The MCS present in the ligand data-set was identified using Multiple MCS programs (Tripod, NIH) and implemented as a template in the template-based method (VLifeMDS, 2010) to obtain optimal topological alignment. We observed six MCSs in the training set (Figure 2) with heterocycle as a common framework. Six-membered rings were the most common MCS shared by all training molecules and selected as the template. Fused heterocycles and heterocycles with connected carbon linkers were the fewer MCSs present in at least three or four molecules, respectively. We have chosen molecular topology-based alignment by creating homogenous molecules i.e. all non-carbon atoms in the molecules are substituted with the carbon atoms, then called ‘Carbon’ molecules, for the following two reasons: (i) Homogenous carbon molecules (resembling carbon skeletons) can be applied to achieve maximum topology-based alignment and (ii) the use of the selected MCS group as seeds for systematic exhaustive alignment method (VLifeMDS, 2010) to ensure optimal overlay. This alignment method also reduces the conformational space required by atom-to-atom matching technique (Supplementary Figure 1) (Dror et al., 2004; Wolber et al., 2006). Besides homogeneity of carbon topology, we introduced additional conditions to create carbon molecules to reduce chemical complexities.

Creating carbon molecules for molecular topology based alignment

Besides the replacement of non-carbon atoms by carbon atoms, the chemical complexities in the carbon molecules can be reduced by retaining the core structure or scaffold and deleting (i) fragments attached to scaffold

Figure 2. The MCS set from molecular data-set. The numbers beneath the MCS represent its frequency in the training (A), carbon training (molecules developed using carbon substituted method; (B) test and (C) sets, respectively.
except connected to another core group(s) or heterocycle(s) and (ii) highly branched carbon chains emanating from scaffold or even from an aliphatic chain(s). The carbon training molecules constituted seven MCSs (Figure 2) with structurally variable structures in comparison to the original MCS set. Noticeably, these conditions substantially reduced the chemical space and enhanced the share of MCS in the molecules. For example, fused heterocycle (4-hydroxy-octahydro-2H-1-benzopyran-2-one) from the original MCS set was observed in only four training molecules, whereas its related fused heterocycle MCS (decahydronaphthalene) devoid of non-carbon atoms are present in nine training molecules thereby assuring optimal molecular topology alignment.

Similarly, six-membered ring MCS was noticed in all 14 test molecules. The structures of training and test set carbon molecules are shown in Figures 3 and 4, respectively. The user may retain the non-carbon atoms (e.g. B, N, O, P, S, etc.) only if the MCS share such atomic configuration in the entire training set.

The alignment revealed that chemical complexities are large around the aligned six-membered MCS (Figure 5) and span the pharmacophore space. Molecules with greater chemical groups were found to be overlaid with the MCS-connected groups, while terminal groups which lack alignment were externally placed. It can be well distinguished that regions densely populated with chemical groups may encompass pharmacophore space constituting various CPH configurations (Batten et al., 1999) that can be applied to develop pharmacophore similarity-based quantitative activity hypothesis. The overall overlay is acceptable to majority of compounds which secured RMSD below 2 Å with few exceptions (Supplementary Table 1).

**Enumeration of atom connectivity of aligned MCS**

The atom connectivity of aligned MCS was enumerated to cluster molecules in bins which will assist in calculating pharmacophore similarity matrix. The atom labels

![Figure 3. The structures of training molecules developed using carbon substituted method.](image-url)
corresponding to the six-membered MCS ring were initially recognized and labeled conventionally as C1, C2, ..., C6. The chemical groups connected to the MCS group can be clustered based on the atom connectivity details (Table 1) viz. bin 1, bin 2, ..., bin 6. Noticeably, the connected groups can be a fragment (F; connected with one atom of MCS) and/or another fused heterocycle (H; connected with two atoms of MCS). The molecules were clustered based on atom pair connectivity data (Table 2) to increase bin chemical complexities, which in turn increase the counts of perceived pharmacophore features. A molecule with an aliphatic chain (fragment) connected at C3 position to MCS group can be clustered in C3–C4 atom pair bin. Similarly, a molecule with heterocycle fused at C3 and C4 positions of MCS group can also be grouped under C3–C4 bin (Table 2). Care should be taken to ensure that each bin must contain at least one active molecule with variable moderately actives and inactives. The most active molecule (molecule with top pIC50, Bin 5,6 – Tr10, Bin 6,1 – Tr3, Bin 1,2 – Tr11, Bin 3,4 – Tr8 and Bin 4,5 – Tr14) in each bin was selected as ‘reference’ molecule to compute pharmacophore similarity matrix and other computations.

**Enrichment of carbon molecules**

The carbon molecules in aligned coordinate positions were enriched by substituting chemical atoms and groups present in the original data-set which represents the mirror image of aligned carbon molecules (Figure 6) and used for further analysis. The chemical groups are abundant around MCS centroid group, whereas terminal groups are externally aligned due to its highly branched nature of molecules.

**Pharmacophore feature perception on aligned data-set**

The pharmacophore features present in the molecules were detected using VLife MolSign module (VLifeMDS, 2010) without affecting the alignment (Supplementary Table 2). This perception step graphically depicted the pharmacophore space of the MCS-based aligned data-set (Figure 6). The shared MCS contributes AroC (aromatic) feature spanned by densely populated variable counts of the H-bond acceptor (HAc) and donor (HDr) features which are further extended by AroC (MCS connected heterocycle) and AlaC (MCS linked aliphatic chains), respectively. It can be studied that molecules with high chemical diversities increased the counts of encoded features. Te3 has 9 HAc, 14 HDr, 1 AroC, 23 AlaC and 3 PosC summed to 50 perceived features (Supplementary Table 2).

**Development of pharmacophore similarity matrix**

Various pharmacophore features present in the data-set were categorized based on its type viz. HAc, HDr, PosC,
NegC, AroC, and AlaC, respectively, to develop pharmacophore similarity matrix. We have chosen pharmacophore feature-based Tanimoto coefficient as a similarity measure (Barnard, Downs, & Willett, 1998) and used reference molecule of each bin to calculate the similarity with the rest of the clustered molecule. The similarity measure was calculated for six different pharmacophore feature types as follows.

\[
\text{Pharmacophore feature based Tanimoto coefficient} = \frac{F_{\text{common}}}{F_{\text{ref}} + F_{\text{mol}} - F_{\text{common}}} \tag{1}
\]

Total pharmacophore similarity (TPS)

\[
= \sum_{i=1}^{6} \text{(pharmacophore feature based Tanimoto coefficient)} \tag{2}
\]

where \( F \) – features, \( F_{\text{ref}} \) – features of reference molecule, \( F_{\text{mol}} \) – features of non-reference molecule, \( F_{\text{common}} \) – features that are common to reference and non-reference molecules, and \( i \) – feature type.

This similarity measure was estimated in two different ways with respect to the consideration of only bin reference molecule (Table 3) and across reference set (Table 4). The pharmacophore similarity with bin reference molecule signifies the feature similarities across its types. Pharmacophoric similar molecules can be identified by the total pharmacophore feature-based Tanimoto coefficient or total pharmacophore similarity (TPS) metric (high TPS = most similar molecules). For example, Tr12 of Bin 1,2 (TPS = 3.667, Table 3) is pharmacophorically similar to Tr11 (reference molecule of Bin 1,2).

**Figure 5.** The molecular topology based alignment of the carbon molecular data-set. The central core group represents aligned MCS.

**Table 1.** Details on atom connectivity with respect to MCS structure.

| Molecules | Atom(s) connectivity | Connectivity type a |
|-----------|----------------------|---------------------|
| Tr1       | 5,6                  | H                   |
| Tr2       | 5,6                  | H                   |
| Tr3       | 6,1                  | H                   |
| Tr4       | 1                    | F                   |
| Tr5       | 3,4                  | H                   |
| Tr6       | 4                    | F                   |
| Tr7       | 4                    | F                   |
| Tr8       | 4                    | F                   |
| Tr9       | 5,6                  | H                   |
| Tr10      | 5,6                  | H                   |
| Tr11      | 1                    | F                   |
| Tr12      | 1                    | F                   |
| Tr13      | 2                    | F                   |
| Tr14      | 4,5                  | H                   |
| Tr15      | 5,6                  | H                   |
| Tr16      | 5,6                  | H                   |
| Tr17      | 5,6                  | H                   |
| Tr18      | 5                    | F                   |
| Tr19      | 6,1                  | H                   |
| Tr20      | 5,6                  | H                   |
| Tr21      | 1,2                  | H                   |
| Tr22      | 6                    | F                   |
| Tr23      | 4                    | F                   |
| Tr24      | 3                    | F                   |
| Tr25      | 6,1                  | H                   |
| Tr26      | 3,4                  | H                   |
| Te1       | 5                    | F                   |
| Te2       | 5                    | F                   |
| Te3       | 6                    | F                   |
| Te4       | 1                    | F                   |
| Te5       | 1                    | F                   |
| Te6       | 1                    | F                   |
| Te7       | 3,6                  | F                   |
| Te8       | 3,6                  | F                   |
| Te9       | 3,6                  | F                   |
| Te10      | 2                    | F                   |
| Te11      | Scaffold             | –                   |
| Te12      | 6,1                  | H                   |
| Te13      | 6,1                  | H                   |
| Te14      | Scaffold             | –                   |

*aF – fragment, H – heterocycle*
and also shares an identical carbon skeleton (Figure 3). TPS computed across reference set (Table 4) represents the feature contribution rate of each clustered molecule with respect to the reference set. In this case, molecules with the higher TPS indicate strong pharmacophore similarity with the reference set and possess prominent features that are encoded by most actives. For example, Tr21 (Table 4) is very similar to Tr10, Tr3, and partially, Tr11, Tr8, and Tr14, as it shares similar hydroxyl and double-bonded oxygen groups. Further, the TPS calculated with bin reference and its set was used to calculate fitness score to assist pharmacophore model evaluations. The hierarchical clustering of training and test molecules is discussed in Supplementary text and shown graphically in Supplementary Figure 2.

Consensus features and relative index for reference molecules

Fitness calculations which seek to recognize accurate and inaccurate molecular alignments from its fitness score describe the dimension of mapped pharmacophore space and identify molecules that are solely aligned on the basis of the MCS centroid group. Initially, the consensus features encoded by the reference set were recognized by counting the minimum and maximum features in each type (Table 5). For example, the reference molecules viz. Tr10, Tr3, Tr11, Tr8, and Tr14 have HAc feature counts of 6, 5, 3, 1, and 2, respectively, among which 2 can be selected as the consensus HAc feature due to its majority. Additionally, a feature type can be discarded from the consensus features set if the feature type was absent in the majority of the compounds with few exceptions. NegC feature is present only in Tr8 and Tr14 among the reference sets, and therefore, discarded from calculations. This discarding criterion should be withheld if the most active molecule among the data-set constitutes the respective feature type because the chemical group mapped by this feature will govern biological response (Dror et al., 2004). The most active molecules, Tr12 (pIC$_{50}$ = 6.432) and Tr13 (6.222), do not constitute the NegC feature type and are discarded in the consensus feature set.

A relative measure called ‘Relative index’ based on the feature counts of consensus and reference molecule was computed to account the extent of availability of a specific feature type in a reference molecule. For example, Tr10 relative index for AroC is 1.667 as it constitutes 5 AroC features as opposed to 3 AroC consensus feature (relative index > 1 indicates the availability of additional features compared to consensus) (Table 5) which shows the possibility of mapping more than 3 AroC features in pharmacophore space.

Table 2. The clustered molecules in each bin based on MCS atom connectivity.

| Bin | Clustered molecules$^a$ |
|-----|-------------------------|
| 5,6 | Tr1, Tr2, Tr9, **Tr10**, Tr20 |
| 6,1 | Tr3, Tr19, Tr22, Tr25, Te3, Te12, Te13 |
| 1,2 | Tr4, **Tr11**, Tr12, Tr13, Tr21, Te4, Te5, Te6, Te10 |
| 3,4 | Tr5, Tr6, Tr7, **Tr8**, Tr23, Tr24, Tr26, Te7, Te8Te9, Te11, Te14 |
| 4,5 | **Tr14**, Tr15, Tr16, Tr17, Tr18, Te1, Te2 |

$^a$The reference molecules of bin are shown in bold face.
The relative index of reference set across each feature type was summed together to interpret the complexities in chemical groups and feature counts in reference molecules. This parameter was called ‘Relative score (RS).’ In our case, Tr10 (RS = 15.667) and Tr8 (2.5) constitute large and small feature complexities (Table 5), respectively. It is also used in the calculations of fitness score.

Relative index

\[
= \text{Pharmacophore feature of the reference molecule/consensus pharmacophore feature}
\]  

(3)

Relative score (RS) = \( \sum \text{Relative index}_{i=1 \text{ to } 6} \)  

(4)

### Table 3. Pharmacophore feature based Tanimoto coefficient and total pharmacophore-similarity (TPS) of non-reference molecules.

| Molecules | HAc  | HDr  | AroC | AlaC | NegC | PosC | Total TPS   |
|-----------|------|------|------|------|------|------|-------------|
| Tr1       | .3333| .125 | .4   | .6   | 0    | 0    | 1.4583     |
| Tr2       | .5   | .25  | .4   | .75  | 0    | 0    | 1.9        |
| Tr4       | .6666| .25  | .5   | 0    | 0    | 0    | 1.4166     |
| Tr5       | .25  | 0    | .6666| 0    | 0    | 0    | .9166      |
| Tr6       | .3333| 1    | .6666| 0    | 0    | 0    | 2          |
| Tr7       | .3333| .5   | .6666| 0    | 0    | 0    | 1.5        |
| Tr9       | .8333| .25  | .8   | .75  | 0    | 0    | 2.6333     |
| Tr12      | .6666| 1    | 1    | 1    | 0    | 0    | 3.6666     |
| Tr13      | 1    | .2   | .25  | 1    | 0    | 0    | 2.45       |
| Tr15      | .5   | 0    | 1    | 0    | 0    | 0    | 1.5        |
| Tr16      | .6666| 0    | .6666| 0    | 0    | 0    | 1.3333     |
| Tr17      | .5   | 0    | .3333| 0    | 0    | 0    | .8333      |
| Tr18      | .5   | 1    | .6666| 0    | 0    | 0    | 2.1666     |
| Tr19      | .4   | .6666| .3333| 0    | 0    | 0    | 1.6        |
| Tr20      | .5   | .125 | .6   | .75  | 0    | 0    | 1.975      |
| Tr21      | .6   | .5   | .75  | .25  | 0    | 0    | 2.1        |
| Tr22      | .4   | .75  | .6666| 0    | 0    | 0    | 1.8166     |
| Tr23      | .3333| 1    | .6666| 0    | 0    | 0    | 2          |
| Tr24      | 0    | 1    | .6666| 0    | .5   | 0    | 2.1666     |
| Tr25      | 0    | 0    | .6666| 0    | 0    | 0    | .6666      |
| Tr26      | 1    | .5   | .3333| 0    | 1    | 0    | 2.8333     |
| Te1       | .5   | 1    | 1    | 0    | 0    | 0    | 2.5        |
| Te2       | .4   | .5   | .75  | 0    | 0    | 0    | 1.65       |
| Te3       | .8333| .5   | 1    | 0    | .3571| 0    | 2.6904     |
| Te4       | 1    | .2   | .5   | .5   | 0    | 0    | 2.2        |
| Te5       | 0    | .25  | .5   | 0    | 0    | 0    | .75        |
| Te6       | 0    | .25  | .5   | 0    | 0    | 0    | .75        |
| Te7       | .125 | .0714|.3333 | 0    | 0    | 0    | .5297      |
| Te8       | .1111|.0714|.3333 | 0    | 0    | 0    | .5158      |
| Te9       | .125 | .125 | .3333| 0    | 0    | 0    | .5833      |
| Te10      | .3333| 1    | .5   | 1    | 0    | 0    | 2.8333     |
| Te11      | .5   | .2   | .3333| 0    | .5   | 0    | 1.5333     |
| Te12      | .2   | .6   | .3333| .4   | 0    | 0    | 1.5333     |
| Te13      | .2   | .3333| .3333| .4   | 0    | 0    | 1.2666     |
| Te14      | .5   | 1    | .3333| 0    | 1    | 0    | 2.8333     |

*Pharmacophore features: HAc – H bond acceptor, HDr – H bond donor, AroC – aromatic group, AlaC – aliphatic group, PosC – positive ionizable center and NegC – negative ionizable center; Pharmacophore feature based Tanimoto coefficient are calculated using Equation (1).

**Fitness score**

Similar to TPS metric, the fitness score was estimated by considering bin-specific reference molecule and reference set as separate entities. Fitness score based on bin-specific reference molecule considers molecular TPS calculated with respect to bin-reference molecule and its reference molecule RS, to identify the alignment of accurate and inaccurate molecular pairs (Supplementary Table 3). An overlay of the reference set with its predicted pharmacophore features is shown in Figure 7. The linear chain of seven aromatic moieties is the core component for alignment spread by HDr features from hydroxylation pattern.

Fitness score for clustered molecules

\[
= \text{Molecule TPS/Bin specific reference molecule RS}
\]  

(5)
The fitness score for clustered molecules was assessed to examine the extent of low and high pharmacophore features engaged in alignment and to identify molecular pairs having low and high TPS and RS, respectively (Figure 8). Tr25 secured low TPS and Tr12 constituted high TPS, whereas Tr8 has a low RS and Tr10 possesses high RS. Noticeably, a molecule with low TPS has fewer features and vice versa and molecules with low RS will possess only essential features (due to less chemical complexity) and vice versa. A trade-off between the perceived features and its chemical complexities can be studied using this fitness score. The overlay of low TPS and RS (Tr25:Tr8) revealed that the MCS group and its features were solely matched and aligned (Figure 8(A)). A better topological alignment was observed for low TPS and high RS pairs (Tr25:Tr10) (Figure 8(B)). Tr25 was aligned with one of the Tr10 terminal fused heterocycle groups. The analyses of molecular pairs with high TPS showed that the alignment with low RS reference molecule (Tr8) (Figure 8(C)) exhibited inaccurate overlay; whereas alignment with high RS molecule (Tr10) (Figure 8(D)) showed the best possible alignment among the data-set. It is evident that optimal molecular pair alignment can be obtained by optimizing the two parameters, TPS and RS above the average of the fitness score. This fitness score also facilitates in the identification of accurate and inaccurate alignments by witnessing its low and high scores, respectively.

Similarly, the fitness score calculated against the reference molecular set in an alignment independent manner (Supplementary Table 4) can be used to map features in pharmacophore space and provides list of molecules that do not embed in the pharmacophore similarity space defined by training molecules. A histogram depicting the distribution of fitness scores (Supplementary Figure 3) was used to examine the external test molecules which do not get placed in the pharmacophore space and achieved poor alignment with the reference molecule under consideration. The lower and upper bounds were defined by the minimum and maximum fitness scores of training molecules against each reference molecule. The minimum and maximum fitness scores computed against Tr10 reference molecule were Tr25 (fitness score = .025) and Tr9 (.168) (Supplementary Table 4) and defined as lower and upper bounds for external test set pursuing alignment with Tr10 as reference molecule. Noticeably, the molecular overlays of poorly aligned pairs, viz. Te7: Tr8, Te8:Tr8, Te10:Tr8, Te13:Tr8, Te7:Tr14, and Te8: Tr14, are graphically illustrated in Figure 9. The MCS was only shared in the alignment of bisdistamycin Te7 with Tr7 (Figure 9(A)) and Tr8 (Figure 9(B)) molecules. A similar mode of alignment was observed for the other bisdistamycin, Te8 with Tr8 (Figure 9(C)) and Tr14 (Figure 9(D)), respectively. Te10 (Figure 9(E)) and Te13 (Figure 9(F)) only shared MCS groups in common with the reference molecule Tr8. These fitness scores cannot be used to predict outliers of the quantitative hypothesis as the trained SVM model samples the pharmacophore space without 3D geometric placement of features and consider orientation independent distance-based pharmacophoric descriptors.

**Applicability domain**

It is very essential to select data-sets with structural diversities and activities span (Dror et al., 2004; Guner, 2000;
Li et al., 2000) and the existence of similarity between training and test sets in pharmacophore modeling (Zhang, Golbraikh, Oloff, Kohn, & Tropsha, 2006). We selected applicability domain to determine this similarity threshold using a modified version of Euclidean distance-based domain (Zhang et al., 2006) based on molecular TPS (a similarity measure) calculated using MCS-based alignment. The domain was calculated by taking the average of TPS from the training set. Subsequently, training molecules with TPSs less than the initial average (1.854) were selected to calculate new average \( <d> \) (1.304) and standard deviation \( \sigma \) (0.371). This new average \( <d> \) is called applicability domain average threshold (Equation (6)) where \( Z \) is an empirical cut-off factor with a default value of 0.5. In addition, an upper bound for pharmacophore similarity space was defined by selecting the most active molecule Tr22 (molecule with high TPS, 1.817). The external test molecular TPS occupied within this domain was considered as suitable molecules for modeling or otherwise treated as outliers of respective bins whose reference molecule may not participate in better alignment (Figure 10). An external test molecule can have better alignment and share pharmacophore similarity with more than one bin-reference molecule due to the flexible chemical structures and therefore, plotted outside the domain threshold. Te1 (2-benzyl-9,10-dihydroanthracene), Te10 ((cyclopenta-2,4-dien-1ylmethyl)benzene), and Te14 (benzene) constitute simple chemical frameworks and share pharmacophore similarity and alignment with more than one reference molecule i.e. these molecules can be placed in other bins too. Te3 and Te4 have complex structures and possess excess features in addition to consensus features and therefore participate as members in other bins.

\[
APD = <d> + \sigma Z
\]

**Quantitative pharmacophore hypothesis and activity predictions**

Distance-based pharmacophoric descriptors were determined by calculating the distance between the perceived feature positions and MCS-aligned centroid positions. Calculations on HIV-1 IN data-set yielded 58 descriptor
variables from which inconsistent descriptors were removed. Finally, a descriptor set of nine features (HAcS – 3, HDrs – 2, AroCs – 2, snd AlaCs – 2) were chosen as independent variables for quantitative SVM modeling (Figure 11, Supplementary Table 5). Mustata et al., 2004 CPH model developed on this data-set obtained a hypothesis with four features (1 hydrophobe, 2 H-bond acceptors and 1 H-bond donor) as the top scored and statistically significant model. We defined a restriction that the external test set must constitute features (common) within the distance limits of training molecules. A quantitative pharmacophore hypothesis which utilizes SVM model trained by training set descriptors will treat an external test molecule as outliers due to non-compliance of distance limits.

We divided the original data-set of 26 training molecules into 21 training and 5 validation sets, respectively. The external test set of 14 molecules was used to predict the biological activities from the trained quantitative SVM model. A grid search was used to identify two optimal anova kernel-based models (Table 6) which secured a significant $r^2$ and $q^2$, respectively. The predictions of external test set from trained models yielded an acceptable range of activities reflected by its $r^2_{ext}$ parameter (Table 7). Mustata

![Figure 8](image_url). The alignment of molecular pairs with low and high TPS and RS estimated from fitness score. Low TPS – Tr25 (A), high TPS – Tr12 (B), low RS – Tr8 (C) and high RS – Tr10 (D). Molecules with less complexity are highlighted in yellow stick model.
et al., 2004 introduced an uncertainty parameter (set to 3) for compound activities to represent the ratio between the upper range of biological activity for the compound and the actual activity (Guner, 2000). Applying this criterion in SVM models, compounds Te12 (residual = 2.894 as per SVM model 1) and Te14 (3.031) should have a partial match with the perceived features geometrically mapped in the pharmacophore space.

The examination of \( r^2_{\text{ext}} \) and activity residuals suggested that SVM model 1 was superior in its performance and predictions compared to the model 2. Distinguishably, the activity residual (error as per Catalyst/HypoGen program (Catalyst, 2000)) for training set in CPH model (Mustata et al., 2004) is capricious in comparison to the observation of significant \( r^2 \) and \( q^2 \) values proposed by our SVM models. Thus,

---

**Figure 9.** The poor molecular aligned pairs predicted by fitness score of external test molecules computed against all reference molecules. This fitness score enables graphical understanding of insignificant alignments between training and external test sets. Molecules with less complexity are highlighted in yellow stick model.

Table 6. Statistics of quantitative pharmacophore SVM models and its parameters.

| Type of molecules | Parameters | SVM model 1 | SVM model 2 |
|-------------------|-----------|-------------|-------------|
| Training          | \( r^2 \) | .9966       | .9971       |
| Validation        | \( q^2 \) | .6696       | .6674       |
| External test set | \( r^2_{\text{ext}} \) | .5366       | .2033       |
| Data-set          | Kernel    | Anova       | Anova       |
|                   | MSE       | .009544     | .010027     |
|                   | \( C \)   | 1           | 1           |
|                   | \( \varepsilon \) | .1          | .1          |
|                   | \( \gamma \) | .1          | .2          |
|                   | \( k_d \) | 5           | 2           |
there is a dependency to focus both on the geometric fit of pharmacophore features, feature selections and its relative count to generate hypothesis. These limitations can be overcome by incorporating partial match criteria to define pharmacophore with consistent descriptors.

Benefits of quantifying pharmacophore space

The chemical complexities and the abundance of encoded pharmacophore features of the diverse data-set paves way for problems associated with conformational sampling, molecular alignment, common features elucidation, and modeling (Dror et al., 2004; Van Drie, 2003; Wermuth & Langer, 1993). The advantages of pharmacophore space and CPH models are described in Supplementary text with an illustration (Supplementary Figure 4). Assuming a low energy conformer as bioactive, a MCS-based alignment can be effectively performed by selecting the common MCS shared by the training set. The carbon-substituted method can also be used to reduce the complexity of the molecule by carefully selecting the MCS and its connected groups to increase the effectiveness of optimal topology alignment. In addition, distance-based pharmacophoric descriptors eases the distance flexibilities among the feature set as it considers the MCS-aligned position as the centroid to calculate distance with consistent features instead of individual feature interdistances. Fitness score and applicability domain provide an overview of geometric alignments and possible outliers. New feature can be effectively incorporated in the proposed pharmacophore space, while the addition of new feature requires developing new CPH. It also evades the possibility of prioritizing only top-scoring CPHs. Finally, a non-linear relationship between biological activities and pharmacophoric descriptor set using SVM modeling is statistically appealing than the CPH model to model predictions.
Table 7. Actual and predicted activities from SVM quantitative hypothesis models.

| Molecules     | Training set | Validation set | External test set |
|---------------|--------------|----------------|------------------|
|               | Actual pIC50 | SVM model 1    | SVM model 2      |
|               |              |                |                  |
| Te13          | 6.432        | 6.222          | 6.332            |
| Te14          | 6.013        | 3.922          | 3.922            |
| Te15          | 3.492        | 4.555          | 3.592            |
| Te16          | 3.492        | 4.010          | 3.592            |
| Te17          | 3.488        | 4.010          | 3.588            |
| Te18          | 3.914        | 4.013          | 4.010            |
| Te19          | 4            | 4.100          | 4.100            |
| Te20          | 3.699        | 3.799          | 3.799            |
| Te21          | 3.533        | 3.633          | 3.633            |
| Te22          | 3.649        | 3.749          | 3.749            |
| Te23          | 3.745        | 3.644          | 3.645            |
| Te24          | 3.798        | 3.798          | 3.798            |
| Te25          | 4.883        | 4.474          |                  |
| Tr1           | 5.678        | 4.883          | 4.474            |
| Tr2           | 4.744        | 4.100          | 4.153            |
| Tr6           | 6.222        | 4.309          | 4.453            |
| Tr7           | 7.398        | 7.053          | 7.329            |
| Tr11          | 7.480        | 3.972          | 3.923            |
| Tr12          | 7.824        | 4.847          | 4.952            |
| Tr13          | 6.155        | 5.455          | 5.271            |
| Tr14          | 5.013        | 3.783          | 4.547            |
| Tr15          | 4.632        | 4.568          | 4.725            |
| Tr16          | 6.638        | 4.826          | 4.819            |
| Tr17          | 6.356        | 4.800          | 4.857            |
| Tr18          | 7.046        | 5.0594         | 4.416            |
| Tr19          | 7            | 5.02           | 4.171            |
| Tr20          | 7.155        | 5.031          | 4.379            |
| Tr21          | 6.886        | 5.132          | 4.887            |
| Tr22          | 6.585        | 5.026          | 4.896            |
| Tr23          | 6.495        | 3.601          | 3.786            |
| Tr24          | 6.444        | 3.674          | 3.878            |
| Tr25          | 5.824        | 2.793          | 3.305            |

*Outliers of SVM quantitative hypothesis models.

Conclusions

A quantitative pharmacophore hypothesis was developed using pharmacophore similarity concept and proposed a ligand-based computational procedure to quantitatively map pharmacophore space by enhancing flexibilities in geometric constraints associated with molecular alignment and pharmacophore perception. A MCS-based topology alignment was performed for the molecular data-set using homogenous carbon molecules and followed by replacement of the original molecules. The pharmacophore features were perceived in alignment mode to calculate pharmacophore similarity matrix which helps in understanding the pharmacophore similarity among the data-set. Fitness score was estimated based on pharmacophore similarity measure to identify molecular pairs with better and inaccurate alignments. Further, an applicability domain was also discussed to predict outliers of bin-specific TPS. A descriptor set of nine features spanning the pharmacophore space was trained in quantitative SVM model to predict the activities of the external test set. The outliers were predicted based on fitness score, and applicability domain and SVM model predictions were discussed. It is convenient to map pharmacophore space instead of the top-scoring CPH models tailored by preselected parameters, especially for diverse data-set to develop quantitative pharmacophore hypothesis models.

Supplementary material

The supplementary material for this paper is available online at http://dx.doi.10.1080/07391102.2014.908142.

Acknowledgments

Sivakumar Prasanth Kumar acknowledges support from the Department of Science and Technology, New Delhi, India, as INSPIRE Senior Research Fellowship. We thank Prakash C. Jha for his excellent technical contributions and helpful discussions.

References

Barnard, J. M., Downs, G. M., & Willett, P. (1998). Chemical similarity searching. Journal of Chemical Information and Computer Sciences; 38, 983–996.

Batten, L., Bohun, C. S., Cheng, K., Doman, T., Drew, J., Edwards, R., Wood, K. L. (1999). Classification of chemical compound pharmacophore structures. Proceedings of the third PIMS Industrial Problem Solving Workshop, University of Victoria. Retrieved from http://faculty.uoit.ca/bohon/pubs/searle.pdf

Bonachéra, F., & Horvath, D. (2008). Fuzzy tricentric pharmacophore fingerprints. 2. Application of topological fuzzy pharmacophore triplets in quantitative structure–activity relationships. Journal of Chemical Information and Modeling, 48, 409–425.

Catalyst. (2000), Catalyst (ver 4.6). San Diego, CA: Accelrys Inc. Retrieved from http://www.accelrys.com

Cato, S. J. (2000). Exploring pharmacophores with Chem-X. In O. F. Guner (Ed.), Pharmacophore perception, development, and use in drug design (pp. 107–126). La Jolla, CA: International University Line.

Chang, C. C., & Lin, C. J. (2001). LIBSVM: A library for support vector machines. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf

Clement, O. A., & Mehl, A. T. (2000). HipHop: Pharmacophore based on multiple common-feature alignment. In O. F. Guner (Ed.), Pharmacophore perception, development, and use in drug design (pp. 69–84). La Jolla, CA: International University Line.

Cramer, R. D., Patterson, D. E., & Bunce, J. D. (1988). Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. Journal of the American Chemical Society, 110, 5959–5967.
Dixon, S. L., Smodyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., & Friesner, R. A. (2006). PHASE: A new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening. Journal of Computer-Aided Molecular Design, 20, 647–671.

Dror, O., Shulman-Peleg, A., Nussinov, R., & Wolfson, H. J. (2004). Predicting molecular interactions in silico: I. A guide to pharmacophore identification and its applications to drug design. Current Medicinal Chemistry, 11, 71–90.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., ... Kollman, P. (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. Journal of Computational Chemistry, 24, 1999–2012.

Gardiner, E. J. (2013). Graph applications in chemoinformatics and structural bioinformatics. In Information Resources Management Association, USA (Ed.), Bioinformatics: Concepts, methodologies, tools, and applications (pp. 1126–1157). Hershey, PA: IGI Global.

Goldgur, Y., Craigie, R., Cohen, G. H., Fujiwara, T., Yoshinaga, T., Fujishita, T., ... Davies, D. R. (1999). Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design. Proceedings of the National Academy of Sciences, 96, 13040–13043.

Guner, O. F. (Ed.). (2000). Pharmacophore perception, development, and use in drug design. La Jolla, CA: International University Line.

Haigh, J. A., Pickup, B. T., Grant, J. A., & Nicholls, A. (2005). Small molecule shape-fingerprints. Journal of Chemical Information and Modeling, 45, 673–684.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Taiwan: Department of Computer Science and Information Engineering, National Taiwan University.

Kirchmair, J., Laggner, C., Wolber, G., & Langer, T. (2005). Comparative analysis of protein-bound ligand conformations with respect to catalyst’s conformational space subsampling algorithms. Journal of Chemical Information and Modeling, 45, 422–430.

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: Structure, Function, and Bioinformatics, 57, 678–683.

Kumar, S. P., Jasrai, Y. T., Pandya, H. A., George, L. B., & Patel, S. K. (2013). Structural insights into the theoretical model of Plasmodium falciparum NADH dehydrogenase and its interaction with artemisinin and derivatives: Towards global health therapeutics. OMICS: A Journal of Integrative Biology, 17, 231–241.

Kumar, S. P., Jasrai, Y. T., Pandya, H. A., & Rawal, R. M. (2013). Pharmacophore-similarity-based QSAR (PS-QSAR) for group-specific biological activity predictions. Journal of Biomolecular Structure and Dynamics. doi:10.1080/07391102.2013.849618.

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 methods: A case study on Plasmodium falciparum Fab enzymes. Journal of Molecular Recognition, 27, 215–229.

Laggner, C., Schieferer, C., Fiechtner, B., Poles, G., Hoffmann, R. D., Glossmann, H., ... Moebius, F. F. (2005). Discovery of high-affinity ligands of signal receptor, ERG2, and emopamil binding protein by pharmacophore modeling and virtual screening. Journal of Medicinal Chemistry, 48, 4754–4764.

Lemmen, C., & Lengauer, T. (2000). Computational methods for the structural alignment of molecules. Journal of Computer-Aided Molecular Design, 14, 215–232.

Li, H., Sutter, J., & Hoffmann, R. (2000). HypoGen: An automated system of generating 3D predictive pharmacophore models. In O. F. Guner (Ed.), pharmacophore perception, development, and use in drug design (pp. 171–189). La Jolla, CA: International University Line.

Mustata, G. I., Brigob, A., & Briggsga, J. M. (2004). HIV-1 integrase pharmacophore model derived from diverse classes of inhibitors. Bioorganic & Medicinal Chemistry Letters, 14, 1447–1454.

Patel, Y., Gillet, V. J., Bravi, G., & Leach, A. R. (2002). A comparison of the pharmacophore identification programs: Catalyst, DISCO and GASP. Journal of Computer-Aided Molecular Design, 16, 653–681.

Purohit, R., Rajasekaran, R., Sudandirarossi, C., George Priya Doss, C., Ramathan, K., & Rao, S. (2008). Studies on flexibility and binding affinity of Asp25 of HIV-1 protease mutants. International Journal of Biological Macromolecules, 42, 386–391.

Purohit, R., Rajendran, V., & Sethumadhavan, R. (2011). Studies on adaptability of binding residues flap region of TMC-114 resistance HIV-1 protease mutants. Journal of Biomolecular Structure and Dynamics, 29, 137–152.

Purohit, R., & Sethumadhavan, R. (2009). Structural basis for the resilience of Darunavir (TMC114) resistance major flap mutations of HIV-1 protease. Interdisciplinary Sciences: Computational Life Sciences, 1, 320–328.

Schneidman-Duhovny, D., Dror, O., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2008). PharmaGist: A webserver for ligand-based pharmacophore detection. Nucleic Acids Research, 36, W223–W228.

SPSS Inc. (2007). SPSS for Windows (16.0). Chicago, IL: IBM Corporation.

Steindl, T., Laggner, C., & Langer, T. (2005). Human rhinovirus 3C protease: Generation of pharmacophore models for peptidic and nonpeptidic inhibitors and their application in virtual screening. Journal of Chemical Information and Modeling, 45, 716–724.

Van Drie, J. H. (2003). Pharmacophore discovery – Lessons learned. Current Opinion in Chemical Biology, 7, 1649–1664.

VLifeMDS: Molecular Design Suite. (2010). Pune: VLife Sciences Technologies. Retrieved from www.vlifesciences.com

Wermuth, C. G., Ganelin, C. R., Lindberg, P., & Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). Annual Reports in Medicinal Chemistry, 33, 385–395.

Wermuth, C. G., & Langer, T. (1993). Pharmacophore identification. In H. Kubinyi (Ed.), 3D QSAR in drug design: Theory, methods and applications (pp. 117–136). Leiden: ESCOM.

Wolfner, G., Dornhofer, A. A., & Langer, T. (2006). Efficient overlay of small organic molecules using 3D pharmacophores. Journal of Computer Aided Molecular Design, 20, 773–788.

Wolfner, G., Seidel, T., Bendix, F., & Langer, T. (2008). Molecule-pharmacophore superpositioning and pattern matching in computational drug design. Drug Discovery Today: 13, 23–29.

Zhang, S., Golbraikh, A., Oloff, S., Kohn, H., & Tropsa, A. (2006). A novel automated lazy learning QSAR (ALL-QSAR) approach: method development, applications, and virtual screening of chemical databases using validated ALL-QSAR models. Journal of Chemical Information and Modeling, 46, 1984–1995.