Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2008) 966–976
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

# Exploring QSTR and toxicophore of hERG $K^+$ channel blockers using GFA and HypoGen techniques

Divita Garg, Tamanna Gandhi, C. Gopi Mohan *

*Centre for Pharmacoinformatics, National Institute of Pharmaceutical Education and Research (NIPER), Sector 67, S.A.S. Nagar 160062, Punjab, India*

Received 26 May 2007; received in revised form 10 August 2007; accepted 10 August 2007
Available online 17 August 2007

**Abstract**

Predictive quantitative structure–toxicity and toxicophore models were developed for a diverse series of hERG $K^+$ channel blockers, acting as anti-arrhythmic agents using QSAR+ module in Cerius2 and HypoGen module in Catalyst software, respectively. The 2D-QSTR analysis has been performed on a dataset of 68 molecules carefully selected from literature for which $IC_{50}$ values measured on hERG $K^+$ channels expressed in mammalian cells lines using the voltage patch clamp assay technique were reported. Their biological data, expressed as $IC_{50}$, spanned from 7.0 nM to 1.4 mM, with 7 orders difference. Several types of descriptors including electrotopological, thermodynamic, ADMET, graph theoretical (topological and information content) were used to derive a quantitative relationship between the channel blockers and its physico-chemical properties. Statistically significant QSTR model was obtained using genetic function approximation methodology, having seven descriptors, with a correlation coefficient ($r^2$) of 0.837, cross-validated correlation coefficient ($q^2$) of 0.776 and predictive correlation coefficient ($r^2_{pred}$) of 0.701, indicating the robustness of the model. Toxicophore model generated using HypoGen module in Catalyst, on these datasets, showed three important features for hERG $K^+$ channel blockers, (i) hydrophobic group (HP), (ii) ring aromatic group (RA) and (iii) hydrogen bond acceptor lipid group (HBAl). The most predictive hypothesis (Hypo 1), consisting of these three features had a best correlation coefficient of 0.820, a low rms deviation of 1.740, and a high cost difference of 113.50, which represents a true correlation and a good predictivity. The hypothesis, Hypo 1 was validated by a test set consisting of 12 molecules and by a cross-validation of 95% confidence level. Accordingly, our 2D-QSTR and toxicophore model has strong predictivity to identify structurally diverse hERG $K^+$ channel blockers with desired biological activity. These models provide a useful framework for understanding binding, and gave structural insight into the specific protein–ligand interactions responsible for affinity, and how one might modify any given structure to mitigate binding.


*Keywords:* QT prolongation; hERG inhibition; 2D-QSTR; Toxicophore; HypoGen

## 1. Introduction

The scrutiny of high attrition rates of clinical candidates led to recognition by the pharmaceutical community that undesirable absorption, distribution, metabolism, elimination and toxicity (ADMET) properties of a molecule are major cause for decline in drug discovery process. The post-marketing scenario shows that toxicity is a major reason for drug withdrawals and 25% of these cases can be attributed to their potential for causing cardiac arrhythmias. Virtually all cases of drug-induced arrhythmias can be attributed to QT prolongation (drug-induced long QT syndrome) which can further be traced to interaction with a particular cardiac ion channel known as human Ether-a-go-go Related Gene (hERG) [1–5]. At present, different *in vivo* and *in vitro* models for QT prolongation causing arrhythmias exist but the predictive power to humans is limited.

The key mechanism for drug-induced QT prolongation is the increased repolarization duration through blockade of outward $K^+$ currents (especially the delayed rectifier repolarizing current, IK) [6]. Because the QT interval reflects the duration of individual action potentials in cardiac myocytes, prolongation of the action potential duration (APD) will result in a prolonged QT interval. Cardiac APD is controlled by a fine balance between inward and outward currents in the plateau

* Corresponding author. Tel.: +91 172 2214 682; fax: +91 172 2214692.
*E-mail addresses:* cmohan@niper.ac.in, cgopimohan@yahoo.com (C. Gopi Mohan).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.08.002

repolarization phase. Since outward $K^+$ currents, especially the delayed rectifier repolarizing current, IK (which is the sum of two kinetically and pharmacologically distinct types of $K^+$ currents: a rapid, IKr, and a slow, IKs, component), play an important role during plateau repolarization and in determining the configuration of the action potential; small changes in conductance can significantly alter the effective refractory period, and hence the action potential duration. In particular, most of the QT prolonging drugs has been shown to inhibit the rapid component of IK, named IKr. Blockade of the hERG, a six transmembrane helices voltage-gated channel $K^+$ channel, is therefore, the most important mechanism through which QT prolonging drugs increase cardiac action potential duration.

hERG mainly comprises of a positively charged S4 helix and a P-loop. Homology modeling [2,3,7], site-directed mutagenesis studies [8], QSAR studies [9–12], and pharmacophore models [13–18] have been reported on the hERG blockers in an attempt to understand the ligand-hERG interactions. These studies have proposed Phe-656 and Tyr-652 of S6 helix to be the key residues involved in hERG-ligand binding. Phe-656 has been proposed to interact with hydrophobic moieties in the ligand by making π-stacking interactions and Tyr-652 makes cation-π interactions. Other key residues include Thr-623 and Ser-624, which modulate the binding potency for a variety of hERG blockers. They interact with the polar tails present in many of the potent blockers [5,15,19]. Also it has been found that the ligands exhibit difference in the channel state (activated or inactivated) they block and this may prevent the application of single pharmacophore model. It has been found that the high-affinity ligands (e.g. dofetilide, cisapride) interact with the inactivated state exhibiting limited voltage-dependent block. On the other hand, low-affinity ligands (e.g. chloroquine, quinidine) interact with the activated state [20] and are characterized by significant voltage-dependent kinetics and steady-state effects [21]. $IC_{50}$ for inhibition of hERG $K^+$ channels expressed in different cell lines is considered as a primary test to study the QT prolonging potential of a molecule [6].

Review of literature shows other *in silico* models for hERG channels developed by different research groups. Computational approaches using different software module such as DISCO in Sybyl [13,14] HypoGen in Catalyst [16,17] and Binary QSAR in MOE [18] have been successfully employed for this purpose. The reported models subsumes by Matyus and colleagues, a five point pharmacophore consisting of hydrogen bond donor, hydrogen bond acceptor, two aromatic rings, and an aliphatic chain (as a hydrophobic group) representing for the most active molecules and a four point pharmacophore developed for the less potent agents [13]; Ekins et al. developed using a training set of 15 molecules, a pharmacophore model in Catalyst, consisting of 4 hydrophobic features and 4 positive ionisable feature. Their proposed distances in the model between the positive center and the four hydrophobes are 5.2, 6.2, 6.8 and 7.5 Å [17]; Cavilli et al. constructed a pharmacophore made up of three aromatic moieties connected through a nitrogen function (tertiary amine) and their distances are 5.2–9.1, 5.7–7.3 and 4.6–7.6 Å, developed using 31 hERG channel blockers. Apart from distances other geometrical parameters namely angles, angle between the planes and height above the planes have also been reported by this group [9]; Du et al. reported using Catalyst 4.9, a pharmacophore model consisting of two aromatic rings, one hydrophobic group and one positive ionisable group along with the geometrical distances developed from 34 molecules [16].

*In silico* modeling techniques offer an important approach towards predicting the potential hERG channel blockers during the early drug development stages. The aim of this work was to develop an integrated quantitative structure–toxicity relationship (QSTR) and toxicophore model for preliminary screening out of cardiotoxic hERG $K^+$ channel blockers. These models provide rich information in the context of virtual screening of relevant libraries and the filtered molecules should be directed for docking analysis in the homology modeled structures of hERG $K^+$ channel at different states. Further the computed ligand bound complex structures can be used to guide synthesis of analogs and later in the clinical stages for developing pharmacological safer drugs with reduced hERG liability.

## 2. Methodology

### 2.1. Quantitative structure–toxicity relationship (QSTR)

Apart from the class IIIA anti-arrhythmics which inhibit hERG ion channels for their therapeutic effects, it has been estimated that approximately 2–3% of all drug prescriptions involve medications that may unintentionally cause the long QT syndrome [6]. Moreover, due to its unique ligand-binding site's shape and hydrophobic nature this channel enjoys interaction with pharmaceuticals of widely varying structure. Thus, it seems reasonable that a large dataset of molecules should be available for this study. However, to study inhibition, hERG channels may be expressed in mammalian or non-mammalian cell lines. The use of the non-mammalian systems leads to a significant underestimation of a drug's potency as a hERG $K^+$ channel blocker, the observed $IC_{50}$ values may be increased by as much as 100 times. The mammalian cell lines on the other hand give a better representation of the behavior of molecules in the human body [9]. Thus, for the purpose of this study, only those drugs were selected from the literature, for which $IC_{50}$ values measured on hERG $K^+$ channels expressed in mammalian cell lines using the voltage patch clamp assay technique were reported. Finally 68 molecules were chosen [22–31] (Fig. A, Supplementary material). The molecules were built with sketch molecule module in molecular modeling package Sybyl7.1 [32] installed on Silicon Graphics Fuel workstation and energy minimized using the semi-empirical AM1 [33] method. The molecules were divided into a training set of 56 molecules (Table 1) and test set of 12 molecules (Table 2), maintaining the structure and activity diversity in both the sets for QSTR model development. The molecular descriptors were calculated using QSAR+ module of Cerius$^2$4.10 [34] software. Before commencing with the development of the 2D-QSTR model, the correlation matrix of about 156 descriptors was calculated and highly correlated

Table 1
Actual experimental $IC_{50}$ ($IC_{50}$ obs) (or $pIC_{50}$ obs) data and predicted $pIC_{50}$ ($pIC_{50}$ pred) of training set molecules based on 2D-QSTR model

| Molecule | $IC_{50}$ obs (μM) | $pIC_{50}$ obs | $pIC_{50}$ pred | Residual[a] |
|---|---|---|---|---|
| Amiodarone | 0.700 | 0.154 | 0.746 | −0.592 |
| Amitryptiline | 10.000 | −1.000 | −0.956 | −0.044 |
| Astemizole | 0.009 | 2.046 | 2.204 | 0.158 |
| Azimilide | 0.560 | 0.252 | −0.074 | 0.326 |
| Bepridil | 0.550 | 0.260 | 0.048 | 0.212 |
| Chloroquine | 2.500 | −0.398 | −0.096 | −0.302 |
| Chlorpheniramine | 1.600 | −0.204 | −0.231 | 0.027 |
| Chlorpromazine | 1.470 | −0.167 | −0.403 | 0.236 |
| Ciprofloxacin | 966.000 | −2.985 | −2.352 | −0.633 |
| Cisapride | 0.007 | 2.155 | 1.858 | 0.297 |
| Clozapine | 0.320 | 0.495 | −0.055 | 0.550 |
| Clozapine-N-oxide | 133.300 | −2.125 | −2.615 | 0.490 |
| Desipramine | 1.390 | −0.143 | −0.859 | 0.716 |
| Diltiazem | 17.300 | −1.238 | −0.749 | −0.489 |
| Diphenhydramine | 30.000 | −1.477 | −0.808 | −0.669 |
| Dofetilide | 0.012 | 1.921 | 2.166 | −0.245 |
| Dolasetron | 5.950 | −0.775 | 0.001 | −0.776 |
| Domperidone | 0.162 | 0.791 | 1.325 | −0.534 |
| Droperidol | 0.032 | 1.495 | 1.737 | −0.242 |
| E-4031 | 0.008 | 2.097 | 1.160 | 0.937 |
| Flecainide | 3.910 | −0.592 | −0.228 | −0.364 |
| Gatifloxacin | 130.000 | −2.114 | −2.286 | 0.172 |
| Granisetron | 3.730 | −0.572 | 0.190 | −0.762 |
| Grepafloxacin | 50.000 | −1.699 | −2.673 | 0.974 |
| Halofentrine | 0.197 | 0.706 | 0.037 | 0.669 |
| Haloperidol | 0.280 | 0.553 | 1.539 | −0.986 |
| Imipramine | 3.400 | −0.531 | −0.678 | 0.147 |
| Lidocaine | 262.900 | −2.420 | −1.402 | −1.018 |
| Loratadine | 0.173 | 0.762 | 0.441 | 0.321 |
| Lumefantrine | 8.130 | −0.910 | 0.445 | −1.355 |
| Mesoridazine | 0.550 | 0.260 | 0.136 | 0.124 |
| Mibefradil | 1.430 | −0.155 | 0.436 | −0.591 |
| Mizolastine | 0.350 | 0.456 | 1.055 | −0.599 |
| Mefloquine | 2.640 | −0.422 | −0.857 | 0.435 |
| Moxifloxacin | 129.000 | −2.111 | −2.090 | −0.021 |
| N-desmethyl-clozapine | 4.490 | −0.652 | −0.176 | −0.476 |
| Ofloxacin | 1420.000 | −3.152 | −2.490 | −0.662 |
| Olanzapine | 6.010 | −0.779 | −0.841 | 0.062 |
| Ondansetron | 0.810 | 0.091 | 0.418 | −0.327 |
| Perhexiline | 7.800 | −0.892 | −0.671 | −0.221 |
| Phenobarbital | 3.000 | −0.477 | −1.483 | 1.006 |
| Phenytoin | 240.000 | −2.380 | −1.825 | −0.555 |
| Pimozide | 0.018 | 1.745 | 1.984 | −0.239 |
| Propafenone | 0.440 | 0.356 | −0.428 | 0.784 |
| Pyrilamine | 1.100 | −0.041 | −0.277 | 0.236 |
| Queitiapine | 5.770 | −0.761 | −0.664 | −0.097 |
| Quinidine | 0.400 | 0.397 | −0.079 | 0.476 |
| Risperidone | 0.167 | 0.777 | 0.448 | 0.329 |
| Sertindole | 0.014 | 1.854 | 1.107 | 0.747 |
| Sparfloxacin | 18.000 | −1.255 | −0.935 | −0.320 |
| Spironolactone | 23.000 | −1.362 | −1.439 | 0.077 |
| Terfenadine | 0.056 | 1.252 | 1.446 | −0.194 |
| Thioridazine | 0.036 | 1.447 | 0.735 | 0.712 |
| Vardenafil | 12.800 | −1.107 | −0.428 | −0.679 |
| Verapamil | 0.143 | 0.845 | 0.265 | 0.580 |
| Ziprasidone | 0.169 | 0.772 | 0.255 | 0.517 |

[a] Residual = ($pIC_{50}$ obs − $pIC_{50}$ pred).

descriptors, with correlation value above 0.7 were removed (Supplementary material, Table A). The remaining descriptors were used for genetic function approximation (GFA) analysis. After analyzing the correlation matrix, the following class of descriptors electrotopological, thermodynamic, ADMET and graph theoretical (topological and information content) were considered for statistical fitting (Table 3).

The electrotopological descriptors are numerical values computed for each atom in a molecule and which encode information about both topological environments of that atom and electronic interactions due to all other atoms in the molecule. The topological relationship is based on the graph distance to each atom. The electronic aspect is based on intrinsic state and perturbation due to intrinsic state differences between atoms in the molecule. The graph-theoretic descriptors base their calculation on representation of molecular structures as graphs, where atoms are represented by vertices and covalent bonds by edges. These descriptors can be categorized into topological and information content descriptor. The former helps in differentiating the molecules according to their size, degree of branching, flexibility, overall shape and views the molecule graphs as connectivity structures. The latter helps in viewing molecule graphs as sources of certain probability distributions to which Shanon's statistical information theory (which measures the uncertainty associated with a random variable) can be applied. All these descriptors are performed on hydrogen suppressed graphs. Thermodynamic descriptors describes energy of molecules and their conversions, while the ADMET class of descriptors help in assessing the absorption, distribution, metabolism, elimination and toxicity attributes of a molecule.

Genetic function approximation developed by Rogers and Hopfinger was employed for statistical analysis to select the relevant descriptors and model them to the channel blocking activity of the molecules. GFA is genetics based method of variable selection which combines Holland's genetic algorithm (GA) with Friedman's multivariate adaptive regression splines (MARS). It works in the following way: first of all a particular number of equations (set at 100 by default in the Cerius2 software) are generated randomly, then pairs of ''parent'' equations are chosen randomly from this set of 100 equations and ''crossover'' operations are performed at random [35]. After some preliminary observations on initial runs, the number of GFA crossovers was set to 20,000 for the present study to obtain a reasonable convergence. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) score, assigned by GFA that resists over fitting and estimates appropriate number of variables in the equation. $r$ is the regression coefficient and is given by the following formula:

$$\mathrm{LOF} = \frac{\mathrm{LSE}}{\{1 - (c + d\,p)/m\}^2}$$

where LSE is the least-squares error, $c$ the number of basis functions in the model, $d$ the smoothing parameter, $p$ the number of descriptors and $m$ is the number of observations in the training set. The smoothing parameter which controls the scoring bias between equations of different sizes was set at a default value of 1.0 and the new term was added with a probability of 50%.

The QSTR model was rigorously evaluated using the leave multiple out approach, Y-randomization test and external test

Table 2
Actual experimental $IC_{50}$ ($IC_{50}$ obs) data (or $pIC_{50}$ obs) and estimated $pIC_{50}$ ($pIC_{50}$ pred) of test set molecules based on 2D-QSTR model and toxicophore model of Hypo 1

| Molecule | $IC_{50}$ obs (μM) | $pIC_{50}$ obs | 2D-QSTR | | Toxicophore | |
|---|---|---|---|---|---|---|
| | | | $pIC_{50}$ pred | Residual[a] | $IC_{50}$ est | Error[b] |
| Acrivastine | 8.000 | −0.90 | −0.93 | 0.03 | 1.5 | −5.33 |
| Amsacrine | 0.204 | 0.69 | 0.67 | 0.02 | 1.4 | 6.86 |
| Cocaine | 4.365 | −0.64 | −0.72 | 0.08 | 1.2 | −3.63 |
| Desmethylastemizole | 1.000 | 0.00 | 0.51 | −0.51 | 0.031 | −32.25 |
| Desmethylcarboxyloratadine | 6.500 | −0.80 | −0.86 | 0.06 | 1.9 | −3.42 |
| Fentanyl | 1.819 | −0.26 | 0.49 | −0.75 | 1.1 | −1.65 |
| Fexofenadine[c] | 21.380 | −1.33 | 0.06 | −1.39 | 0.011 | −1943.63 |
| Ketoconazole | 1.995 | −0.30 | 0.49 | −0.79 | 0.02 | −99.75 |
| Lidoflazine | 0.016 | 1.79 | 1.90 | −0.11 | 0.15 | 9.37 |
| Meperidine | 7.500 | −1.87 | −1.82 | −0.05 | 1.1 | −6.81 |
| Tadalafil | 10.000 | 1.00 | 1.00 | 0.00 | 1.1 | −9.09 |
| Cetrizine | <35 | <−0.47 | −1.694 | Acceptable | 8.7 | ∼−4.02 |

Activity scale: highly active (<1 μM), moderately active (1–20 μM) and weakly active (>20 μM).
[a] Residual = $pIC_{50}$ obs − $pIC_{50}$ pred.
[b] (+) Error = $IC_{50}$ est/$IC_{50}$ obs when actual activity < estimated activity; (−) error = $IC_{50}$ obs/$IC_{50}$ est when actual activity > estimated activity.
[c] Outlier molecule.

set predictions. In the leave multiple out approach, the specified '*n*' number of molecules is removed from the training set, the modified model is prepared and predictions are made for the removed molecules. Y-randomization test confirms whether the model is obtained by chance correlation, and is a true structure–toxicity relationship to validate the adequacy of the training set molecules. The steps followed in randomization test are (a) repeatedly scrambling the activity data in the training set molecules (b) using the randomized data to generate QSTR equations (c) comparing the resulting scores with the score of the original QSTR equation generated with non-randomized data. If the random model's activity prediction is comparable to the original equation, the set of observations is not sufficient to support the model. The predictive properties of the developed model were more rigorously tested by predicting the hERG blocking potency of external test set of 12 molecules.

### 2.2. Toxicophore model generation

As an additional screen for identifying potential hERG channel blockers, a toxicophore model was developed. In the present study the identification of the toxicophore was done using HypoGen module in Catalyst version 4.10 [36]. HypoGen attempts to derive SAR models from a set of molecules for which activity values ($IC_{50}$ or $K_i$) on a given biological target are available. It optimizes hypotheses that are present in the highly active molecules in the training set, but missing among the least active (or inactive) ones, constructs simplest hypotheses that best correlate the activities (estimated versus measured).

The first step in search for toxicophore relies on thorough study of the characteristics of the conformations of low energy for each molecule. Accordingly, the conformational space of each inhibitor was explored by adopting the ''best conformer generation'' option, which is based on CHARMM-like force field implemented using the Poling algorithm in the software [37]. The conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structure with a maximum limit of 255 conformers per molecule. Hence, this search procedure will probably succeed in identifying the best 3D arrangement of chemical functionalities explaining the activity variations among the training set. Typically, HypoGen requires informative training sets of 16–45 molecules with bioactivities spread over at least four orders of magnitude [37–39].

To develop a toxicophore model that incorporated the specific features in the three dimensional spatial arrangement responsible for hERG $K^+$ channel inhibition, 56 molecules of the training set of QSTR, were divided into two clusters, cluster 1 consisting of 43 molecules with a low $IC_{50}$ (less than 10 μM) and cluster 2 consisting of 13 molecules with high $IC_{50}$ (more than 10 μM). Cluster 1 was subsequently used for toxicophore identification, the activity in this set spanned over 4 orders of magnitude. Initially to detect the different possible features that

Table 3
Descriptors used to build 2D-QSTR model

| Type | Descriptors |
|---|---|
| E-state indices | Electro-topological |
| Information content | Information of atomic composition index, information indices based on the A-matrix, information indices based on the D-matrix |
| Topological | Wiener index, Zagreb index, Hosoya index, Kier and Hall molecular connectivity index, Balaban indices |
| ADMET | AlogP98, Receptor_RSA, van der Waals polar surface area |
| Thermodynamic | Log of partition coefficient, log of partition coefficient atom-type value, heat of formation, molar refractivity |

these molecules contain, functional mapping was performed on the most toxic molecules of the training set: cisapride, astemizole and E-4031. This showed the presence of hydrogen bond acceptor (HBA), hydrogen bond acceptor lipid (HBAl), hydrogen bond donor (HBD), hydrophobic group (HP), positive ionisable group (PI) and ring aromatic group (RA). The definitions of HBA and HBAl vary by the inclusion of a $NH_3$ group in the latter. Thus the five features HBAl, HBD, HP, PI and RA were selected for toxicophore hypothesis generation. According to earlier pharmacophore study for hERG channel blockers, the HBD, HBA, RA and HP features were considered to be important chemical features, which supports the present analysis. In order to create the best hypothesis for toxicophore model generation, various HypoGen runs were performed. In the first run for toxicophore identification the default values, i.e. minimum of 0 and maximum of 5 for each selected chemical feature were maintained. Then various other runs were performed by altering the minimum and maximum value for 1 and 5, respectively, for feature hypothesis generation. Ten hypotheses were generated in each run, the best hypothesis was selected on the basis of maximum difference between null cost and total cost, and the lowest root mean square (rms) divergence. Cost values calculated in Catalyst is a criterion to rank the generated hypothesis; null cost represents zero correlation between structure and activity and total cost is the corresponding value calculated for the hypothesis. Large difference between the total cost and null cost (the residual cost) indicates a good correlation between the generated hypothesis and activity.

To appraise the statistical quality of HypoGen model various validation operations apart from the cross-validation based on Fischer's randomization test, applied using the CatScramble program were performed. The latter is similar in principal to the randomization test of QSTR, in which the activity data among the training set molecules is randomized to generate pharmacophore hypotheses, using the same features and parameters that were used to develop the original pharmacophore hypothesis. If the randomized sets generate pharmacophores with similar or better cost values, rms and correlation, then the original pharmacophore can be considered as to be generated by chance, i.e. if the randomized data sets produce a hypothesis with a high correlation value, then the methodology of the pharmacophore generation is flawed. This chemoinformatics approach should be directed towards potential scaffold-based QT liabilities in the course of whole library screens, internal as well as commercial, and real as well as virtual, for identifying ''druggable'' molecules and should be integrated with absorption, distribution, metabolism, elimination and toxicity (ADMET) prediction packages. A further problem in modeling studies on hERG channel blockers (defined as an anti-target) is the recognition of molecules which will prevent binding to the channel. In this context QSTR and structure-based models play the fundamental role of contributing to the discovery of the molecular determinants of hERG binding from the point of view of the ligand and target respectively. For clarity, we have used the terminology QSTR and toxicophore, instead of QSAR and pharmacophore, throughout the paper.

## 3. Results and discussion

### 3.1. QSTR model for hERG $K^+$ channel blockers

The final QSTR model obtained is represented in equation and it consists of seven descriptors and a constant.

$$\begin{aligned} \mathrm{pIC_{50}} = {} & 2.04746 + 0.182271\,\mathrm{kappa3} - 1.61453\,\mathrm{Atype\_O\_57} \\ & + 0.251155\,\mathrm{S\_sNH2} - 1.97268\,\mathrm{JX} \\ & - 0.969258\,\mathrm{Atype\_O\_59} + 0.616081\,\mathrm{ADMET\_PPB} \\ & - 0.063453\,\mathrm{Atype\_H\_46} \end{aligned}$$

$$n = 56, \quad r^2 = 0.837, \quad r^2_{\mathrm{cv}}\,(\mathrm{or}\,q^2) = 0.776,$$
$$\mathrm{BS}r^2 = 0.838 \pm 0.001, \quad \mathrm{LOF} = 0.5, \quad r = 0.925,$$
$$\mathrm{PRESS} = 16.06 \quad \text{and} \quad r^2_{\mathrm{pred}} = 0.701$$

where, '$n$' refers to the number of molecules in the training set, $r^2$ the correlation coefficient, $r^2_{\mathrm{cv}}$ (or $q^2$) is the correlation coefficient obtained by the leave one out (LOO) cross validation procedure. The $\mathrm{pIC_{50}}$ values predicted for the molecules during LOO cross validation are given in Table 1 and the scatter plot of the predicted to observed values are presented in Fig. 1(a). $\mathrm{BS}r^2$ is the correlation coefficient obtained by the bootstrap validation procedure in which a random sample of '$n$' molecules is iteratively removed from the set and model is generated using the remaining molecules. The model is then used to predict the activity value for the removed molecules. The descriptors appearing in the model is presented in Table 4, which provide some insight into the nature of the binding site and structural requirements for potent hERG channel blockers. Kappa3 a topological descriptor, S_sNH2 an electrotopological descriptor and ADMET_PPB an ADMET descriptor showed positive correlation towards the $\mathrm{pIC_{50}}$ value. The Kier & Hall kappa shape index such as kappa3 are intended to capture the different aspects of model shape. These indices compare the model graph with ''minimal'' and ''maximal'' graphs, where the meaning of ''minimal'' and ''maximal'' depends on the order $n$. For order 3, the counts of paths of length 3 are considered. These topological descriptors do not distinguish between heteroatom. The presence of this descriptor in the model shows the importance of shape in exhibiting hERG $K^+$ channel inhibitory activity. S_sNH2 stands for sum descriptor for nitrogen bonded to two hydrogens and one single bond, and indicated that increased presence of features like nitrogen containing heterocyclic moieties contribute more towards hERG $K^+$ channel inhibition. ADMET_PPB descriptor corresponds to likelihood of binding of the molecule to plasma proteins in the threshold of 90–95%. Its value is 0 if binding is $<$90%, 1 if binding is between 90 and 95% and 2 if the binding is $\geq$95%.

The Balaban index, JX reflects the relative connectivity and effective size of the carbon chain to which multiple methyl groups are attached. Balaban index and hERG $K^+$ channel inhibition holds inverse relationship, i.e. more the number of methyl groups, the lower the activity value. The Atype

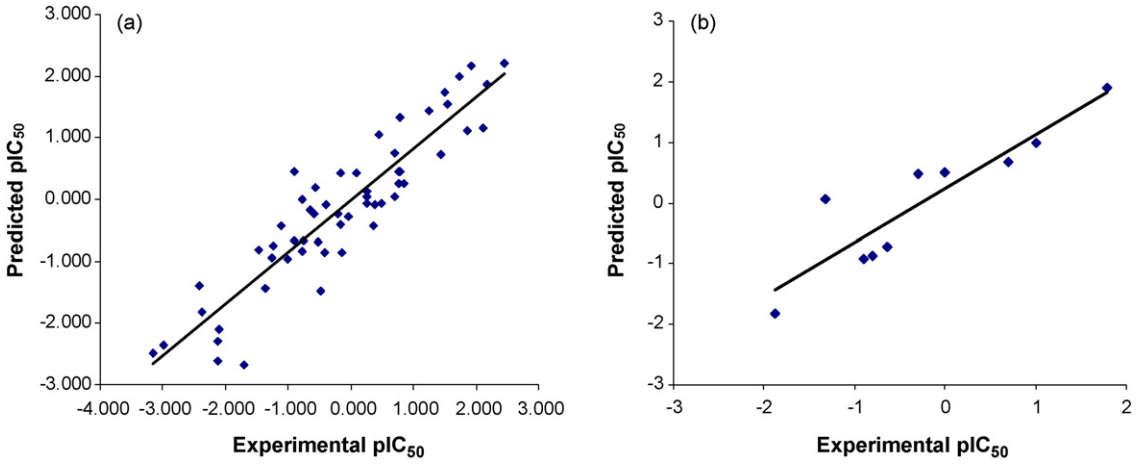


Fig. 1. Scatter plots of predicted versus experimental $pIC_{50}$ values for (a) training set and (b) test set from 2D-QSTR analysis.

descriptors are thermodynamic descriptors defining the presence of that type of atom in the molecule. Various atom type AlogP descriptors can be used to calculate the logP of molecules. In this strategy halogens and hydrogen are classified by the hybridization and oxidation state of the carbon they are bonded to; carbon atoms are classified by their hybridization state and the chemical nature of their neighboring atoms. Atype_H_46 is one of the atom type AlogP descriptors that appeared in the model, in this the ''H'' refers to hydrogen attached to $Csp^0_3$, where 0 represent the formal oxidation number while $sp_3$ denote the hybridization states. The negative slope of this descriptor in the equation explains that hERG $K^+$ channel inhibitory activity decreases with increase in the hydrophobicity associated with this hydrogen. Other atom type descriptors, Atype_O_57 and Atype_O_59 also showed a negative contribution to activity. Here the ''O'' in Atype_O_57 represent oxygen in phenol, enol, carboxyl OH and for Atype_O_59 refers to oxygen in Al–O–Al (Al = aliphatic) [40]. Negative correlation of thermodynamic descriptors (A-type) and topological balaban index (JX) towards $pIC_{50}$ signifies that presence of functional features like electronegative group in the ligand decreases the affinity for hERG $K^+$ channels.

### 3.2. Validation of the 2D-QSTR model

Cross-validation operation of the QSTR results showed a high correlation coefficient obtained by the LOO procedure, $r^2_{cv}$ (or $q^2$) of 0.776, and a comparable PRESS to the original model (Supplementary material, Table B), which is statistically significant. Furthermore, the randomization test, conducted by performing 99 random trials (Supplementary material, Table C) showed a large difference of about 0.7 in the $r$ values for non-random and random trials, with the observed deviation of random values being only 0.09, suggesting that the selected training set was adequate to support the developed QSTR model. Also, none of the random $r$ values were greater than the non-random values. In the external set prediction the residual values for 11 out of the 12 molecules were found to be within the acceptable range of 1. In addition, the $r^2_{pred}$ for the test set molecules was 0.701, which is promising. The scatter plot of experimental to predicted $pIC_{50}$ for test set molecules is presented in Fig. 1(b). Overall, the developed 2D-QSTR model is robust and was found satisfactory for not only predicting the activity of new compounds and its metabolites, but also for explaining the important chemical regions in the molecules in a quantitative manner.

### 3.3. Pharmacophore model for hERG $K^+$ channel blockers

Using HypoGen module different hypothesis were generated in order to identify the best toxicophore model for hERG $K^+$ channel blockers. In the first run for toxicophore identification (with default values for each selected feature) a hypothesis (Hypo 1) with three features, one each of RA, HBAl and HP

Table 4
Descriptors appearing in the 2D-QSTR model

| Label | Type | Definition |
|---|---|---|
| S_sNH2 | E-state indices | Sum descriptor for nitrogen bonded to two hydrogens and one single bond |
| JX | Topological | Balaban index |
| Kappa-3 | | Third order Kier's shape index |
| ADMET_PPB | ADMET | Plasma protein binding tendency |
| Atype_O_57 | Thermodynamics | Atom type O in phenol, enol, carboxyl OH |
| Atype_O_59 | | Atom type O in Al–O–Al |
| Atype_H_46 | | Atom type H attached to $Csp^3_0$ |

Table 5
Information of statistical significance and predictive power presented in cost values along with its residual cost, rms and correlation coefficient of six different hypotheses (Hypo 1–Hypo 6) generated using HypoGen module in Catalyst

| Hypothesis | Features with min 1 to max 5 occurrence | Features with min 0 to max 5 occurrence | Features in the generated hypothesis | Total cost | Residual cost | Rms deviation | Correlation ($r$) |
|---|---|---|---|---|---|---|---|
| Hypo 1 | | HP, HBAl, PI, RA, HBD | HP, HBAl, RA | 208.772 | 113.502 | 1.738 | 0.819 |
| Hypo 2 | HP, HBAl, PI, RA | HBD | HP, HBAl, PI, RA | 221.814 | 100.46 | 1.955 | 0.764 |
| Hypo 3 | HP, PI, RA, HBD | HBAl | HP, PI, RA, HBD | 230.286 | 91.988 | 2.079 | 0.728 |
| Hypo 4 | HBAl, PI, RA, HBD | HP | HBAl, PI, RA, HBD | 229.028 | 93.246 | 2.063 | 0.732 |
| Hypo 5 | HP, HBAl, RA, HBD | PI | HP, HBAl, RA, HBD | 222.392 | 99.882 | 1.974 | 0.759 |
| Hypo 6 | HP, HBAl, PI, HBD | RA | HP, HBAl, PI, HBD | 236.236 | 86.038 | 2.128 | 0.712 |

Null cost of Hypo 1 to Hypo 6 hypothesis is 322.274.
Residual cost = (null cost − total cost).

was obtained. The total cost of Hypo 1 was 208.77 and null cost 322.27. The residual value 113.50 between the null and total costs implies that this hypothesis is not a chance correlation. The statistical results of the hypothesis generated have been summarized in Table 5. On this hypothesis of all the molecules—100% mapped HP function, 91% mapped HBAl function and 72% mapped the RA function. Out of the 44 molecules in the training set, 28 molecules could map over Hypo 1 without any feature miss as shown in Table 6. These included 22 of the 25 molecules whose observed $IC_{50}$ value was less than or equal to 1 μM which indicates a very high affinity channel blockers, whereas the low affinity molecules did not map completely. This emphasizes that the hypothesis could distinguish the highly toxic molecules from less toxic ones. To investigate other feature combinations, further five runs (Hypo 2–Hypo 6) were performed by modifying the default values for four of the features to a minimum occurrence of 1 and a maximum occurrence of 5, leaving the default values for one feature sequentially so that each feature had this combination at least once. This highlighted the importance of each of the five chemical features in bringing about hERG $K^+$ channel inhibition. The results of each run are summarized in Table 5. It was noticeable that in all these runs (Hypo 2–Hypo 6), four feature hypotheses were generated; the fifth parameter for which the occurrence was set at a minimum of zero, failed to appear. The residual cost calculated as a difference between total and null cost is highest for Hypo 1 followed by Hypo 2, Hypo 5, Hypo 4, Hypo 3 and Hypo 6 (Table 5). Analysis of the percentage mappings of molecules over individual features suggests that HP is the most significant feature followed by HBAl, RA and PI respectively, and is presented in Table 6. HBD has very poor percentage mapping on these datasets, indicating it to be of least significance for hERG inhibition.

Complete mapping of molecules without any feature miss over the Hypo 2–Hypo 6 was less than 50%, against 63% for Hypo 1, indicating that Hypo 1 had the highest significance of identifying toxic molecules. Mapping of three most toxic molecules (a) cisapride, (b) astemizole and (c) E-4031) over the best hypothesis, Hypo 1 is presented in Fig. 2. Further support to the better predictability of Hypo 1 is given by the highest difference between the null and total cost values, lowest rms deviation value with significant correlation coefficient ($r$) (Table 5).

### 3.4. Validation of toxicophore model

The main aim of quantitative toxicophore is to identify the potentially toxic new structures and predict their activity value correctly. To verify if Hypo 1 could fulfill this aim, test set analysis on 12 molecules, initially selected during 2D-QSTR model development was performed. All molecules in the test set were built and minimized like all molecules in the training set and the structural data are shown in Fig. A (Supplementary material). In this study all molecules were classified by their activity as highly active (<1 μM), moderately active (1–20 μM) and weakly active (>20 μM) respectively. The test set molecules were mapped onto Hypo 1, the best toxicophore hypothesis and the observed activity versus estimated activity are shown in Table 2. Out of the 12 test set molecules, 8 had an error value up to 10 representing not more than 1 order difference between the observed and estimated values, showing

Table 6
Feature mapping of training set on complete hypotheses without any feature miss and on individual features in the hypotheses

| Hypothesis | %Molecules mapping without any feature miss | %Molecules mapping the individual feature | | | | |
|---|---|---|---|---|---|---|
| | | RA | HBAl | HBD | PI | HP |
| Hypo 1 | 63 | 72 | 91 | – | – | 100 |
| Hypo 2 | 46 | 82 | 75 | – | 88 | 86 |
| Hypo 3 | 2 | 52 | – | 11 | 96 | 98 |
| Hypo 4 | 7 | 84 | 82 | 16 | 96 | – |
| Hypo 5 | 25 | 50 | 84 | 52 | – | 100 |
| Hypo 6 | 11 | – | 82 | 18 | 93 | 96 |

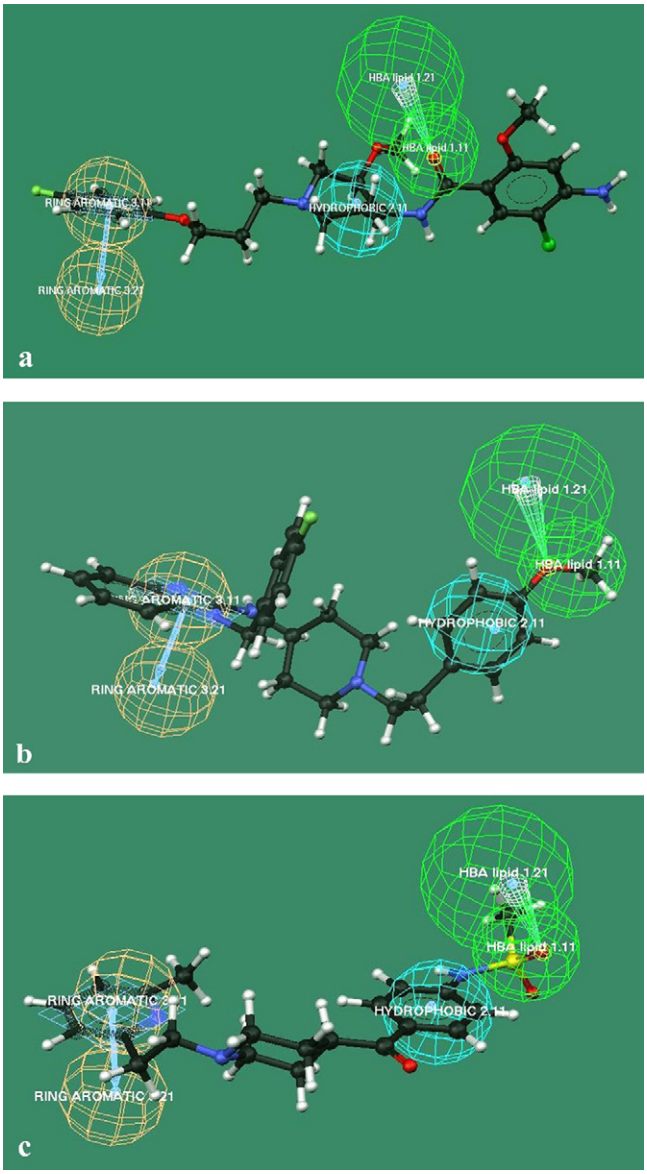


Fig. 2. Best scoring HypoGen generated toxicophore model. Hypo 1, is mapped to the most toxic molecules (a) (cisapride, $IC_{50}$ = 0.007 μM), (b) (astemizole, $IC_{50}$ = 0.009 μM) and (c) (E-4031, $IC_{50}$ = 0.008 μM). Toxicophoric features are color-coded map: orange, ring aromatic (RA); cyan, hydrophobic (HP); and green, hydrogen bond acceptor lipid (HBAl). Distance between toxicophoric features: HP-HBAl: 4.5 Å; RA-HP: 11.4 Å; RA-HBAl: 11.8 Å respectively.

good correlation between observed and estimated activity. In this case, all highly active molecules (amsacrine and lidoflazine) were predicted correctly. Among eight moderately active molecules (acrivastine, cocaine, desmethylastemizole, desmethylcarboxyloratadine, fentanyl, ketoconazole, meperidine and tadalafil), the molecule desmethylastemizole and ketoconazole were overestimated as highly active. Also, the molecule fexofenadine is overestimated as false negative among the three weakly active molecules fexofenadine, tadalafil and cetrizine. Thus among the twelve test set molecules with the exception of the desmethylastemizole, ketoconazole and fexofenadine molecule, were predicted correctly or better than there actual activity (Table 2). The

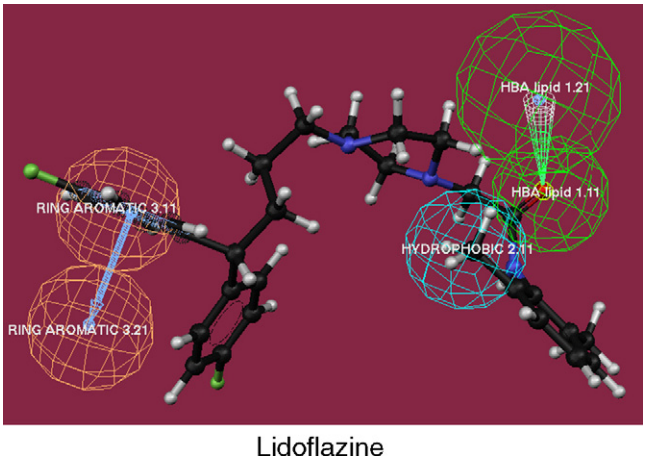


Fig. 3. Best scoring HypoGen generated toxicophore. Hypo 1, is mapped to the most toxic molecule: Lidoflazine, $IC_{50}$ = 0.016 μM, in the test set. Toxicophoric features are color-coded map: orange, ring aromatic (RA); cyan, hydrophobic (HP); and green, hydrogen bond acceptor lipid (HBAl).

most toxic molecule (Lidoflazine, $IC_{50}$ = 0.016 μM) in the test set, was selected to show the mapping of this molecule on Hypo 1 represented in Fig. 3. Lidoflazine molecule chemical features fits the ring aromatic, the hydrophobic and hydrogen bond acceptor lipid group, but does not fit the second aromatic ring.

Another approach to assess the quality of HypoGen hypothesis is to apply the CatScramble program based on Fischer's randomization test. Main aim of this analysis is to check whether there is a strong correlation between the chemical structure and the biological activity. Using CatScramble program, cross-validation test was performed with 95% confidence level for 19 trial runs. The result of the CatScramble indicated that randomization produced hypotheses with low predictivity, and is listed in Table 7. Out of the 19 trial runs, only three, had a correlation value more than 0.6, but the rms deviation of these trial runs was very high with very less residual cost, which is not desirable for a good hypothesis. Hence unscrambled hypothesis (Hypo 1) was much superior showing significant statistical results, than that of the scrambled hypothesis (Table 7). This cross-validation also provided strong support on the initial toxicophore Hypo 1.

As an additional proof of HypoGen hypothesis generator to discriminate between high and low toxic molecules and to determine the goodness of Hypo 1, an additional hypothesis (Hypo 7) was generated for the training set having activity >10 μM, (low toxicity set). The hypothesis generated had two HBAl features and one RA feature. None of the molecules from the complete data set mapped over this hypothesis (Hypo 7) completely. The difference in the chemical features of Hypo 1 and Hypo 7, and poor mapping over Hypo 7 indicate the ability of the hypothesis generator to distinguish between the high and low toxic molecules, and emphasizes further the goodness of Hypo 1.

Published pharmacophore models of hERG channel blockers typically have three important chemical features, i.e. basic nitrogen center flanked by aromatic and hydrophobic groups [13–18]. Our hERG toxicophore model also have the same

Table 7
Results from cross-validation using CatScramble in Catalyst

| Trial no. | Total cost | Null cost | Residual cost | rms deviation | Correlation (*r*) |
|---|---|---|---|---|---|
| Results for unscrambled (Hypo 1) | 208.77 | 322.27 | 113.50 | 1.74 | 0.82 |
| Results for scrambled | | | | | |
| 1 | 289.15 | 322.27 | 33.13 | 2.66 | 0.48 |
| 2 | 266.78 | 322.27 | 55.49 | 2.39 | 0.62 |
| 3 | 262.61 | 322.27 | 59.67 | 2.37 | 0.62 |
| 4 | 302.64 | 322.27 | 19.63 | 2.74 | 0.44 |
| 5 | 265.45 | 322.27 | 56.82 | 2.42 | 0.60 |
| 6 | 277.59 | 322.27 | 44.69 | 2.57 | 0.53 |
| 7 | 269.89 | 322.27 | 52.39 | 2.45 | 0.59 |
| 8 | 276.42 | 322.27 | 45.86 | 2.50 | 0.57 |
| 9 | 299.67 | 322.27 | 22.61 | 2.75 | 0.43 |
| 10 | 286.09 | 322.27 | 36.18 | 2.60 | 0.51 |
| 11 | 251.70 | 322.27 | 70.57 | 2.31 | 0.47 |
| 12 | 278.63 | 322.27 | 43.64 | 2.52 | 0.56 |
| 13 | 273.65 | 322.27 | 48.63 | 2.48 | 0.57 |
| 14 | 309.14 | 322.27 | 13.14 | 2.79 | 0.39 |
| 15 | 283.64 | 322.27 | 38.63 | 2.60 | 0.52 |
| 16 | 259.83 | 322.27 | 62.45 | 2.35 | 0.63 |
| 17 | 297.63 | 322.27 | 24.65 | 2.69 | 0.46 |
| 18 | 277.18 | 322.27 | 45.10 | 2.50 | 0.56 |
| 19 | 283.49 | 322.27 | 38.79 | 2.61 | 0.51 |

chemical features, as the second best hypothesis (Hypo 2), showing its predictive power to be at par with those reported by other research groups. In addition, Hypo 2 feature mapping on the training data set also showed that the presence of PI to be most prominent followed by HP, RA and HBAl (Tables 5 and 6). The total cost of Hypo 2 was 221.81 and null cost 322.27. The residual value 100.46 between the null and total costs implies that this hypothesis is significant. On this hypothesis percentage molecules mapping the feature are - 88% mapped PI function, 86% mapped HP function, 82% mapped the RA function and 75% mapped HBAl function. Out of the 44 molecules in the training set, 20 molecules could map over Hypo 2 without any feature miss as shown in Table 6. In addition, from feature mapping studies carried out on various generated hypothesis Hypo 2 to Hypo 6, the function PI, mainly mapped by the nitrogen atom shown highest percentage of mapping despite less residual cost, correlation coefficient and high rms and is presented in Tables 5 and 6. Its presence, in addition to the HP, RA and HBAl chemical features could increase the hERG binding propensity of the molecule. This observation, i.e. presence of PI chemical feature is also supported by the positive contribution of S_sNH2 descriptor in the QSTR model. The HBAl chemical feature which support our best hypothesis Hypo 1 having 91% feature mapping do also support the positive contribution of S_sNH2 descriptor. In addition, HBA chemical feature is less significant in our model and the negative correlation towards activity of atom type descriptors, Atype_O_57 (oxygen in phenol, enol, carboxyl OH) and Atype_O_59 (oxygen in Al–O–Al (Al = aliphatic)) appearing in our QSTR model do support this hypothesis.

However, statistically best hypothesis, Hypo 1 in the present toxicophore analysis showed the presence of hydrophobic and aromatic chemical features, but it lacked the basic nitrogen centre. The absence of chemical feature corresponding to basic nitrogen in Hypo 1 can be justified with reference to uncharged hERG actives like mizolastine, ondansetron, ketoconazole, which though actives lack the basic nitrogen centre [18]. Some blockers may contain aromatic ring in place of basic nitrogen centre, the former may participate in favorable π-stacking interactions in lieu of a cation interaction with Tyr-652. Admittedly, pharmacophore for 194 potent uncharged hERG channel blockers was reported lacking the basic nitrogen, which fully support the present Hypo 1 analysis [18]. These mapping features of known hERG blockers onto the proposed pharmacophore can be used as guidance for chemists focused on eliminating hERG liability in an uncharged hERG active chemical series. The fact that both hypothesis (Hypo 1 and Hypo 2) consist of multiple hydrophobes could be a consequence of the generally long flexible molecules in the present data sets. These observations also clearly suggest that there may be multiple binding interactions within the potassium channel and as said earlier depends upon the potency of the inhibitor as well as the activation potential of the hERG $K^+$ channel. Addressing this problem may require multiple pharmacophores or platform-specific models that could detect subtle structural differences.

However, it is important to mention that our 2D-QSTR model is able to correctly predict the inhibition of parent molecules and its metabolites. For example, it could distinguish thioridazine and its metabolite mesoridazine; clozapine and its metabolite clozapine N-oxide and N-desmethyl-clozapine; astemizole and its metabolite desmethylastemizole; loratadine and two of its metabolites (Tables 1 and 2). Furthermore, the developed QSTR and pharmacophore model can distinguish potent inhibitors of hERG, i.e. thioridazine, cisapride, lidoflazine and sertindole from its weaker inhibitors, i.e. meperidine, phenytoin and clozapine-N-oxide, thereby enabling us to rank order molecules, which may be valuable in early drug discovery

(Tables 1 and 2). The present QSTR and pharmacophore model is thus robust in the sense that it has the capacity to distinguish the inhibition between parent molecules and its metabolites as well as strong and weak inhibitors of diverse structural classes required for hERG channel blockade. Unfortunately, our two models (QSTR and pharmacophore) failed to predict correctly the activity for fexofenadine, an outlier, and which is an active metabolite of terfenadine (Table 2). The reason could be the minor difference in chemical features (addition of COOH group in fexofenadine) in these two molecules.

The lessons learned from the present analysis is that mapping the chemical space of known hERG blockers has shed some light on the structural aspects of ligands consistent with the channel blocker profile. Also we have noticed that large number of diverse class of hERG channel blockers both basic and neutral have highly flexible linkers and connects various molecular fragments. Due to the high flexibility of the ligands, the bioactive conformation adopted seems to be compatible with the large binding pore region in the channel, viz., ligand rigidification seems to be a proper solution for overcoming the channel complexity.

Also, much information exists for non-blockers than blockers of hERG channels. The ongoing medicinal chemistry optimization program have direct impact on this. The drive toward better computational models which we followed in the present analysis may receive a guidance from increasing availability of high-throughput hERG channel assays (planar patch or patch clamp assay techniques).

The model based on literature data therefore provides a potentially valuable tool for discovery chemistry as future molecules may be synthesized that are less likely to inhibit hERG based on information provided by this dual approach. The scope of our two models is surprising in light of the complexity of the hERG channel. Possibilities for additional binding features not covered by our models are entirely plausible given the complexity of the potassium channel, its state-dependent conformational changes, and its interactions with the membrane and regulatory proteins. Furthermore, factors such as membrane permeation required for accessing the intra-cellular lumen of the channel and possible shifts in local $pK_a$ due to changes in membrane potential further complicate the problem of predicting hERG channel inhibitory activity. The ability to locate and describe these alternative channel active sites is likely to come from an concerted effort of *in vitro*, *in vivo* and *in silico* experiments such as radioligand competition assays, site directed mutagenesis, homology or comparative modeling and ultimately biocrystallography.

The present results can be used to flag commercial libraries overly enriched with predicted hERG platform specific channel blockers. In addition to this the *in vitro* and *in vivo* models are very labor-intensive, costly and not widely available, and is therefore generally being performed at the post-clinical stages. Nevertheless, it is hoped that our 2D-QSTR and toxicophore models can be used as a preliminary guidance for explaining hERG channel liabilities in early lead candidates as well as for designing out properties that promote hERG channel blocking in such molecules.

## 4. Conclusions

The challenge for drug development programs is to minimize the hERG $K^+$ channel blockade, without discarding therapeutically novel drugs needed to treat a number of diseases. We present a 2D-QSTR and toxicophore model for LQTS-inducing drugs, for the hERG $K^+$ channel blocking activity. The QSTR model was developed from an organized list of QT-prolonging drugs selected following the criterion of the homogeneity of both the mechanism causing the QT prolongation and the *in vitro* assay system in transfected mammalian cells. In an attempt to extend the understanding of the physicochemical characteristics underlying the biological activity of the drugs investigated and to provide a predictive tool for the hERG $K^+$ channel blocking activity, we developed a toxicophore model based on the same set of molecules used for the generation of the 2D-QSTR and tested through an external set of molecules. The toxicophore model generated has three important chemical features, i.e. hydrophobic, ring aromatic and hydrogen bond acceptor lipid, as reported earlier by other research groups. The results we obtained allow one to draw a consistent description of the molecular features associated with the QT-prolonging ability and eventually to assess quantitatively the HERG $K^+$ channel blocking potential of new molecules. The 2D-QSTR model presented here offers an acceptable level of predictivity, but cannot yet be proposed as a rapid and efficient *in silico* tool for the early identification of the LQTS-inducing activity. Both QSTR and toxicophore models were thoroughly validated using cross validation, randomization and external test set prediction methodologies. In addition, our models can predict the activity of molecules, not included in model generation and can distinguish between metabolites, toxic and less toxic molecules. It is conceivable that our *in silico* approach offers an efficient screening method for the QT-prolonging potential and can be used in integration with existing *in vitro* methods to increase overall predictivity at the early stages of drugs development. These models therefore provides understanding of important structural features of molecules, a valuable guidance for drug discovery chemistry, and future molecules may be synthesized that are less likely to inhibit hERG $K^+$ channel blockers based on the information provided by these models. This information would reduce the likelihood of developing drugs that could lead to a life-threatening arrhythmias, a surrogate marker for cardiotoxicity.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.08.002.

## References

[1] A.M. Aronov, B.B. Goldman, A model for identifying HERG $K^+$ channel blockers, Bioorg. Med. Chem. 12 (2004) 2307–2315.
[2] R.A. Pearlstein, R.J. Vaz, J. Kang, X.L. Chen, M. Preobrazhenskaya, A.E. Shchekotikhin, A.M. Korolev, L.N. Lysenkova, O.V. Miroshnikova, J. Hendrix, D. Rampe, Characterization of HERG potassium channel

inhibition using CoMSiA 3D QSAR and homology modeling approaches, Bioorg. Med. Chem. Lett. 13 (2003) 1829–1835.
[3] J.S. Mitcheson, J. Chen, M. Lin, C. Culberson, M.C. Sanguinetti, A structural basis for drug-induced long QT syndrome, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 12329–12333.
[4] B.J. Zunkler, Human ether-a-go-go-related (HERG) gene and ATP-sensitive potassium channels as targets for adverse drug effects, Pharmacol. Ther. 112 (2006) 12–37.
[5] M.C. Sanguinetti, M. Tristani-Firouzi, hERG potassium channels and cardiac arrhythmia, Nature 440 (2006) 463–469.
[6] M. Recanatini, E. Poluzzi, M. Masetti, A. Cavalli, F. De Ponti, QT prolongation through hERG K(+) channel blockade: current knowledge and strategies for the early prediction during drug development, Med. Res. Rev. 25 (2005) 133–166.
[7] R. Rajamani, B.A. Tounge, J. Li, C.H. Reynolds, A two-state homology model of the hERG K+ channel: application to ligand binding, Bioorg. Med. Chem. Lett. 15 (2005) 1737–1741.
[8] Q. Wang, N.E. Bowles, J.A. Towbin, The molecular basis of long QT syndrome and prospects for therapy, Mol. Med. Today 4 (1998) 382–388.
[9] A. Cavalli, E. Poluzzi, F. De Ponti, M. Recanatini, Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K(+) channel blockers, J. Med. Chem. 45 (2002) 3844–3853.
[10] G. Cianchetta, Y. Li, J. Kang, D. Rampe, A. Fravolini, G. Cruciani, R.J. Vaz, Predictive models for hERG potassium channel blockers, Bioorg. Med. Chem. Lett. 15 (2005) 3637–3642.
[11] A. Coi, I. Massarelli, L. Murgia, M. Saraceno, V. Calderone, A.M. Bianucci, Prediction of hERG potassium channel affinity by the CODESSA approach, Bioorg. Med. Chem. 14 (2006) 3153–3159.
[12] M. Song, M. Clark, Development and evaluation of an *in silico* model for hERG binding, J. Chem. Inf. Model 46 (2006) 392–400.
[13] P. Matyus, A.P. Borosy, A. Varro, J.G. Papp, D. Barlocco, G. Cignarella, Development of pharmacophores for inhibitors of the rapid component of the cardiac delayed rectifier potassium current, Int. J. Quant. Chem. 69 (1998) 21–30.
[14] A.P. Borosy, K. Keseru, I. Penzes, P. Matyus, 3D QSAR study of class I antiarrhythmics, J. Mol. Struct. 503 (2000) 113–129.
[15] H. Choe, K.H. Nah, S.N. Lee, H.S. Lee, S.H. Jo, C.H. Leem, Y.J. Jang, A novel hypothesis for the binding mode of HERG channel blockers, Biochem. Biophys. Res. Commun. 344 (2006) 72–78.
[16] L.P. Du, K.C. Tsai, M.Y. Li, Q.D. You, L. Xia, The pharmacophore hypotheses of I(Kr) potassium channel blockers: novel class III antiarrhythmic agents, Bioorg. Med. Chem. Lett. 14 (2004) 4771–4777.
[17] S. Ekins, W.J. Crumb, R.D. Sarazan, J.H. Wikel, S.A. Wrighton, Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channels, J. Pharmacol. Exp. Ther. 301 (2002) 427–434.
[18] A.M. Aronov, Common pharmacophores for uncharged human ether-a-go-go-related gene (hERG) blockers, J. Med. Chem. 49 (2006) 6917–6921.
[19] A.M. Aronov, Predictive *in silico* modeling for hERG channel blockers, Drug Discov. Today 10 (2005) 149–155.
[20] J.A. Sanchez-Capula, E. Salinas-Stefanon, J. Torres-Jacome, D.E. Benavides-Haro, R.A. Navarro-Polanco, Blockade of currents by the antimalarial drug chloroquine in feline ventricular myocytes, J. Pharmacol. Exp. Ther. 297 (2001) 437–445.
[21] E. Ficker, C.A. Obejero-Paz, S. Zhao, A.M. Brown, The binding site for channel blockers that rescue misprocessed human long QT syndrome type 2 ether-a-gogo-related gene (HERG) mutations, J. Biol. Chem. 277 (2002) 4989–4998.
[22] C. Chouabe, M.D. Drici, G. Romey, J. Barhanin, M. Lazdunski, HERG and KvLQT1/IsK, the cardiac K+ channels involved in long QT syndromes, are targets for calcium channel blockers, Mol. Pharmacol. 54 (1998) 695–703.
[23] B.R. Danielsson, K. Lansdell, L. Patmore, T. Tomson, Phenytoin and phenobarbital inhibit human HERG potassium channels, Epilepsy Res. 55 (2003) 147–157.
[24] G.J. Diaz, K. Daniell, S.T. Leitza, R.L. Martin, Z. Su, J.S. McDermott, B.F. Cox, G.A. Gintant, The [3H]dofetilide binding assay is a predictive screening tool for hERG blockade and proarrhythmia: comparison of intact cell and membrane preparations and effects of altering [K+], J. Pharmacol. Toxicol. Methods 50 (2004) 187–199.
[25] T. Maurizio, P. Castaldo, A. Pannaccione, G. Giorgio, L. Annunziato, Human ether-a-gogo related gene (HERG) $K^+$ channels as pharmacological targets, Biochem. Pharmacol. 55 (1998) 1741–1746.
[26] J. Kang, L. Wang, X.L. Chen, D.J. Triggle, D. Rampe, Interactions of a series of fluoroquinolone antibacterial drugs with the human cardiac K+ channel HERG, Mol. Pharmacol. 59 (2001) 122–126.
[27] D. Rampe, M.L. Roy, A. Dennis, A.M. Brown, A mechanism for the proarrhythmic effects of cisapride (Propulsid): high affinity blockade of the human cardiac potassium channel HERG, FEBS Lett. 417 (1997) 28–32.
[28] J.J. Salata, N.K. Jurkiewicz, A.A. Wallace, R.F. Stupienski 3rd, P.J. Guinosso, J.J. Lynch Jr., Cardiac electrophysiological actions of the histamine H1-receptor antagonists astemizole and terfenadine compared with chlorpheniramine and pyrilamine, Circ. Res. 76 (1995) 110–119.
[29] D.J. Snyders, A. Chaudhary, High affinity open channel block by dofetilide of HERG expressed in a human cell line, Mol. Pharmacol. 49 (1996) 949–955.
[30] M. Traebert, B. Dumotier, L. Meister, P. Hoffmann, M. Dominguez-Estevez, W. Suter, Inhibition of hERG K+ currents by antimalarial drugs in stably transfected HEK293 cells, Eur. J. Pharmacol. 484 (2004) 41–48.
[31] Z. Zhou, V.R. Vorperian, Q. Gong, S. Zhang, C.T. January, Block of HERG potassium channels by the antihistamine astemizole and its metabolites desmethylastemizole and norastemizole, J. Cardiovasc. Electrophysiol. 10 (1999) 836–843.
[32] SYBYL7.1, Tripos Inc., St. Louis, Mo 6314 USA, 2003.
[33] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1984) 3902–3909.
[34] Cerius$^2$, Version 4.10, A. San Diego, Inc., CA, USA, 2005.
[35] D. Rogers, A.J. Hopfinger, Application of genetic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[36] Catalyst, Version 4.10, A. San Diego, Inc., CA, USA, 2005.
[37] A. Smellie, S.L. Teig, P. Towbin, Poling: promoting conformational variation, J. Comp. Chem. 16 (1995) 171–187.
[38] Y. Kurogi1, O.F. Güner, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.
[39] R.G. Karki, V.M. Kulkarni, A feature based pharmacophore for Candida albicans MyristoylCoA: protein N-myristoyltransferase inhibitors, Eur. J. Med. Chem. 36 (2001) 147–163.
[40] A.K. Ghose, G.M. Crippen, Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure–activity relationships 2. Modeling dispersive and hydrophobic interactions, J. Chem. Inf. Comput. Sci. 27 (1987) 21–35.