Published in final edited form as:

J Chem Inf Model. 2011 November 28; 51(11): 2948–2960. doi:10.1021/ci200271d.

A Critical Assessment of Combined Ligand-based and Structure-based Approaches to hERG Channel Blocker Modeling

Lei Du-Cuny^{*,‡}, Lu Chen^{*,‡}, and Shuxing Zhang^{*,€,*}

[‡]Integrated Molecular Discovery Laboratory, Department of Experimental Therapeutics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

[€]Molecular Modeling and Structural Biology Core, The University of Texas M. D. Anderson Cancer Center, Houston, Texas

Abstract

Blockade of hERG channel prolongs the duration of the cardiac action potential and is a common reason for drug failure in preclinical safety trials. Therefore, it is of great importance to develop robust *in silico* tools to predict potential hERG blockers in the early stages of drug discovery and development. Herein we described comprehensive approaches to assess the discrimination of hERG-active and -inactive compounds by combining QSAR modeling, pharmacophore analysis, and molecular docking. Our consensus models demonstrated high predictive capacity and improved enrichment, and they could correctly classify 91.8% of 147 hERG blockers from 351 inactives. To further enhance our modeling effort, hERG homology models were constructed and molecular docking studies were conducted, resulting in high correlations ($R^2=0.81$) between predicted and experimental binding affinities. We expect our unique models can be applied to efficient screening for hERG blockades, and our extensive understanding of the hERG-inhibitor interactions will facilitate the rational design of drugs devoid of hERG channel activity and hence with reduced cardiac toxicities.

Introduction

Drug cardiac arrhythmia has become a major safety concern for both pharmaceutical development and health regulatory authorities.¹ In recent years, a number of clinically successful drugs have been withdrawn from the market due to their drug-induced sudden cardiac deaths.² This toxicity is caused by the blockade of the human *Ether-à-go-go* related gene (hERG) channel which leads to the long QT interval, an abnormality in cardiac muscle repolarization. This syndrome has been implicated as a predisposing factor for torsades de points, a polymorphic ventricular tachycardia that can spontaneously degenerates to ventricular fibrillation and causes sudden death³. An experimental assessment of hERG inhibition is time consuming and costly⁴; therefore, it is necessary to develop reliable *in silico* modeling methods to screen the hERG activity of drug candidates during the early stages of drug discovery and development. Furthermore, a detailed structural understanding of hERG channel would help design drugs with reduced hERG activity.

The Kv11.1 potassium (K^+) ion channel encoded by hERG is known for contributing to the electrical activity of the heart. However, the 3D structure is not yet to be available. Under this situation, various modeling studies have been performed to assess and predict potential

*Corresponding author: Shuxing Zhang, Ph.D., The University of Texas M. D. Anderson Cancer Center, Department of Experimental Therapeutics, Unit 1950, 1901 East Rd., Houston, TX 77054, Telephone: (713)-745-2958, Fax: (713) 794-5577, shuzhang@mdanderson.org.

[‡]These authors contribute equally to this work

hERG liability. For instance, ligand-based methods, such as principle component analysis (PCA), artificial neural network (ANN) and support vector machine (SVM) along with 2D molecular properties and structural fingerprints are frequently employed to predict hERG liability^{5–10}. Descriptors with easily interpretable physicochemical meanings were applied by Yoshida et al.⁵ to derive a QSAR model with $r^2 = 0.70$ and $q^2 = 0.67$ for 104 hERG blockers. In a validation study, ANN models were used to classify 93% of 72 non-blocking agents and 71% of 23 hERG channel blockers⁶. Yufeng Tseng's group conducted a series of PLS and SVM classification studies of hERG blockers based on Pubchem Bioassay datasets using 4D-Fingerprints and traditional 2D descriptors^{7, 8}. "Shape Signature" molecular descriptors in conjunction with SVM were applied by Ekins et al., and the binary classification model achieved 69–73% cross-validation accuracy for the training set⁹. Other 2D fragment-based approaches, such as hologram QSAR¹⁰ and substructure-based v-support vector regression¹¹, were utilized without explicit consideration of the 3D binding modes and calculation of the molecular descriptors. Comparative molecular field analysis (CoMFA) studies of hERG channel blockers were also described and the authors concluded that four features are critical to the hERG binding activity¹: a flexible molecule, a central tertiary amine, at least two aromatic moieties, and a polar group on one side of the molecule. Ekins et al.¹² and Aronov et al.^{13, 14} generated 3D pharmacophore models for hERG blockers. Hydrophobic and positive ionizable features were found as important for hERG liability. To further understand the structure-function relationships of hERG channel, its 3D structure was constructed using homology modeling^{15, 16}. Li et al. combined GRIND pharmacophore descriptors with SVM method to predict hERG blockades¹⁷. However, this SVM model only achieved 72% accuracy on their test set. Combined receptor and ligand-based approaches were also used to develop a universal pharmacophore models, providing rapid assessment of drug blocking ability to the hERG channel with high accuracy¹⁸. Unfortunately, many of these studies were conducted by commercial groups and their models/data are not publicly available. These published works either employed relatively small collections of hERG blockades or used the datasets in which the IC_{50} or K_i values are divergent and lack data consistency. In addition, there has been limited work to comprehensively benchmark the performance of different computational methods for predictions of potential hERG inhibitors.

In the present study, we compiled a much larger dataset containing 178 hERG blockers, and this enabled us to develop more robust hERG models with higher statistical significance and better predictive capability. Furthermore, we integrated both structure and ligand-based methods and evaluated their performance. Specifically, we developed variable selection k nearest neighbor (kNN) QSAR and 3D pharmacophore models to predict the hERG inhibition. With QSAR methods, a collection of 2D descriptors was applied to a set of known hERG-active ("actives") and hERG-inactive ("inactives") drugs. For the 3D pharmacophore modeling, we employed an automated 3D pharmacophore ensemble approach. In both cases, classification models were developed to discriminate the actives from inactives in the datasets. This provides robust models for rapidly screening large virtual libraries to identify molecules with potential hERG liabilities. In addition, to probe hERG-inhibitor interactions, 3D homology models of hERG channel protein were constructed for molecular docking studies. A set of 12 hERG blockers were used to validate the structural models based on their known binding affinities and available mutagenesis data^{19–26}. Our results were consistent with experimental studies, and the pIC_{50} values could be reproduced for a wide range of ligands against the hERG channel.

Materials and Methods

Datasets

IC₅₀ values for hERG blockers were collected from previous publications^{1, 5, 11, 27–32}. Some discrepancies were observed between the dataset published by Ekins et al.³² and others^{1, 5, 11, 27–31}. Based on thorough literature investigation, we found data from resources other than Ekins were more trustworthy as they were reported by more than one research group and consistent to our own predictions. In total we collected 147 hERG active compounds as **Data Set I**, and although the 31 compounds from Ekins et al.³² were excluded from model building, they were kept as **Data Set III** (Figure 1) for external testing of our models. In addition, a set of compounds was compiled by extracting orally deliverable drugs from the DrugBank³³, with molecular weights less than 500Da. Compounds in this compilation have very diverse structures and they were for different disease treatment including: anti-HIV, anticancer, anti-obesity, antidepressant, etc. We further investigated literature thoroughly for this chemical dataset, and if any relation between any compound and the hERG protein is indicated, the compound will be excluded. This resulted in 351 compounds forming the presumably hERG-inactive set (**Data Set II**). Furthermore, it is known that a basic nitrogen moiety is one of the major features^{1, 31, 34} for hERG activities. Thereby, we divided **Data Set I** into two subsets: compounds with (**Data Set IV**) and without (**Data Set VI**) a basic moiety. Similarly, the inactives (**Data Set II**) were divided into **Data Set V** and **Data Set VII** with and without the basic nitrogen moiety, respectively. The 3D structures of all compounds in our datasets were generated using the Molecular Operating Environment (MOE¹⁴) software. This includes the elimination of counter ions, addition of hydrogen atoms, and calculation of standard protonation states for ionizable atoms at a physiological pH (pH=7.4). The molecules were then subjected to energy minimization to generate the lowest energy conformation.

QSAR modeling of hERG blockers

The 147 hERG blockers (**Data Set I**) with known hERG inhibition activity (pIC₅₀) were used to build regression models. For the classification model construction, both **Data Set II** (351 hERG-inactive compounds) and **Data Set I** (147 hERG-active compounds) were employed to form the training and test sets. MOE¹⁴ was used to calculate 184 molecular descriptors which were normalized to avoid disproportional weighting. Using the sphere exclusion algorithm as we described previously^{35, 36}, the compounds were split into multiple training and test sets for both regression and classification modeling.

The kNN pattern recognition principle³⁷ and a variable selection procedure were applied to develop models of hERG inhibition. Briefly, a subset of n_{var} (number of selected variables) descriptors was selected randomly. The n_{var} parameter was tuned to obtain the best correlation for regression modeling (q^2 value) or best classification (accuracy value) for classification modeling. The models were optimized by leave-one-out cross validation, where each compound was eliminated from the training set and its hERG activity was predicted as the weighted average activity of the k most similar molecules. The molecular similarity was characterized by the Euclidean distances between compounds in multidimensional descriptor space. A simulated annealing method with Metropolis-like acceptance criteria was used to optimize the selection of variables. A variety of models with q^2 greater than 0.4 (for regression) or accuracy greater than 40% (for classification) were obtained and validated on the test sets to select the predictive models with R^2 values >0.5 or accuracy >50%. To test the model reliability, Y-randomization³⁸ (randomization of response) was conducted and repeated three times for the training set using randomized activities. An applicability domain³⁹ with parameter $Z = 0.5$ was employed to ensure that

only compounds similar to training set compounds were predicted and the reliability was high.

3D Pharmacophore Modeling for the Prediction of hERG Blockers

Classic hERG compounds usually contain a positively charged nitrogen center;⁴⁰ however hERG activity has been observed in ligands lacking the moiety.^{41–43} Therefore, two data sets were prepared for the hERG blockers with and without basic moieties to generate 3D pharmacophore models. For the dataset with basic moieties, eight most active compounds studied by Cavalli et al.¹ with pIC₅₀ values greater than 6.9 were used as actives and 191 drugs (**Data Set V**) as inactives. For the dataset without basic moieties, 10 most active neutral compounds from **Data Set VI** with pIC₅₀ values greater than 5.5, along with 160 drugs (**Data Set VII**) as hERG negatives, were used to generate pharmacophore models.

We used the pharmacophore elucidation method in MOE to generate the pharmacophore queries. The queries were selected according to the following steps: The input molecules were first prepared by assigning protonation states at pH 7.0 and removing counter ions. A collection of lowest energy conformers was generated and aligned in a flexible manner. Then the alignments were scored according to the internal strain and overlap of molecular features, e.g. aromaticity, hydrogen bond donor/acceptor, hydrophobicity, etc.⁴⁴. Pharmacophore elucidation was subsequently performed to generate up to 5-point pharmacophore queries using specified MOE pharmacophore annotation schemes, including Unified⁴⁵, PCHD⁴⁵, and PPCH⁴⁵. The “*query spacing*” was set to 0.8 and the “*query cluster*” was set to 1.25 during the query generation. Each query was then used to search the active molecules (8 for the charged and 10 for the uncharged dataset). Only those queries that lead to a good alignment of common pharmacophore features (overlap score⁴⁴ > 1.5 and coverage values > 50%) were retained. Finally, all queries were analyzed and the best pharmacophore query was identified, according to its coverage of hERG actives and its ability to discriminate between the actives and inactives.

hERG Homology Modeling

Five different bacterial K⁺ channels (KcsA, MthK, KvAP, KirBac1.1, Kv1.2) have their crystal structures resolved. The crystal structure of KcsA is in the closed form, while MthK has a tetramer structure but only part of KvAP, KirBac1.1 and Kv1.2 was resolved in open form. Hence, KcsA from *Streptomyces lividans*⁴⁶ (PDB ID 1BL8) was used as the structural template (22% sequence identity) for homology modeling of hERG in the closed form, and MthK from *Methanothermobacter thermautotrophicus*⁴⁷ (PDB ID: 1LNQ) was used to model hERG in the open form (24% sequence identity). The sequence of hERG was aligned to KcsA and MthK using the software ALIGN⁴⁸ (Supplementary Figure 1). Our alignments were identical to those published ones⁴⁹. These sequence alignments were then used to generate ten homology models with MODELLER 9v6⁵⁰. The top model with the lowest MODELLER energy scores was selected for further structural refinement as follows: The rotameric conformations of Tyr652 and Phe656 were manually adjusted so that they point towards the selectivity filter loops¹⁵ of hERG; The steric conflicts were subsequently removed by refining the side chain conformations of the remaining residues based on the backbone-dependent rotamer library⁵¹ in the Coot software⁵²; Superimposition of the S5 helix of the selected lowest energy hERG models reveals that a bend in the S6 helix at Gly648 permits the structural differences that are observed between the closed and open states. The S6 helices in the closed and open state differentiate themselves about 40 degrees. To derive the intermediate structures of hERG in different partially open states, the S6 helices in the closed form were bent towards (with Gly648 as the hinge point) the open form conformations using Coot⁵² in a step of about 8 degree each time, as done by Rajamani¹⁶. The side chains of these derived partially open states were further refined using Coot.

Molecular Dynamics for Structural Optimization—Molecular dynamics simulations were carried out in AMBER 8⁵³ to further refine the hERG homology models using the generalized Born solvation (GB) model with the FF99 force field⁵⁴. The side chains were minimized with 500 steps of steepest descent methods followed by 500 conjugate gradient steps. Subsequently, the system was gradually heated to 300K and the dynamics time step is 2fs. During the first 1.6ps, the protein backbone coordinates were kept as rigid using harmonic constraints. With each monomer of only 56 residues, hERG pore domain is a relatively simple homo-tetramer and its binding pocket consists of only four helices without any loops. Thus, we performed 500ps MD structural refinement in GB solvation model with constraints removed, as suggested by some previous report⁵⁵.

Molecular Docking—The average structure of the ensembles of MD simulation was extracted for the molecular docking. Mutagenesis data provides important insights for binding pose selections. Herein, we collected 12 compounds along with their binding data change upon the mutation of several critical protein residues (Table 1A) to validate the 3D structure of the hERG models by analyzing the interactions between the hERG channel and its blockers. The program GOLD⁵⁶ was employed for the molecular docking of these 12 compounds with default parameters unless otherwise noted. We found Phe656 is critical for ligand binding⁵⁷ and sits in the middle of the channel. Therefore, the binding pocket was defined as 15Å around the centroid of the four Phe656 residues of the tetramer that forms the pore domain (subunits S5–S6) of the hERG channel. For the selection of docking poses, residues important for ligand binding based on mutagenesis data (Table 1A) were employed to assess the hERG-inhibitor binding.

Evaluation of the hERG Homology Models Using MacroModel and SIMCA—Six models were generated to capture the flexibility of the hERG channel of different open-close conformations. In order to discover which particular models are accurately constructed, each model was evaluated using 12 hERG blockers with known binding affinities (pIC₅₀ values) and mutagenesis data as follows: the selected docking pose of each ligand was minimized within the protein binding pocket using the MacroModel module in Maestro⁵⁸. The OPLS_2005 force field⁵⁹ and the generalized Born/surface area (GB/SA) continuum water model were applied. The minimization was continued until either the energy converged (0.05 convergence criterion) or the maximum 5000 iterations was reached. The protein residues within 8Å of the inhibitors were optimized during the minimization. All other residues were held as rigid. After the ligand-receptor complex was minimized, the ligand was extracted from the complex. The 3D structures of the protein and ligand were then minimized separately using a GB/SA continuum water model to obtain reference energies for their free states. The computed differences (ΔG_{bind}) in the stretch, bend, torsion, improper torsion, van der Waals, and electrostatic energies between the bound (G_{complex}) and free states (G_{receptor} , G_{ligand}) (Equation 1) were used to correlate with the experimental pIC₅₀ values.

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \quad \text{Equation 1}$$

PLS (partial least-square) was used to predict the binding affinity of hERG blockers using the program SIMCA⁶⁰. To this end, the above energetic terms (Table 1B) were used as descriptors ($\Delta\text{Stretch}$, ΔBend , $\Delta\text{Torsion}$, $\Delta\text{Improper Torsion}$, ΔVDW , $\Delta\text{Electrostatic}$) to establish their correlation with pIC₅₀. These descriptors were normalized to the unit variance. Principal component analysis (PCA) was used to detect outliers as follows: new latent variables were calculated by summarizing the information contained in the original variables. The compounds which could not be well explained with the latent variables were classified as outliers. Outliers were identified using the 95% tolerance interval, which is

signified as an ellipse in the PCA loading plot⁶⁰ (Supplementary Figure 2A). The goodness of the fit (q^2) of a PLS model was evaluated with leave-one-out cross-validation. The q^2 value was used as the main criterion for assessing the quality of a given model. Once a model was chosen, it was validated by a Y-randomization test by scrambling the Y values to ensure that the model was not obtained by chance. For a generated model, the influence of every original variable on the matrix Y was estimated using the VIP (variable influence on projection) value⁶⁰ (Supplementary Figure 2C). In a case where a strong correlation between pIC₅₀ and energy descriptors was found, the corresponding model structure can be employed to guide the hERG liability optimization.

Results and Discussion

QSAR regression Models for the Quantitative Prediction of hERG inhibition

Six QSAR models (Supplementary Table 1) were obtained from kNN analyses with q^2 greater than 0.55 for the training set and R^2 greater than 0.55 for the test set compounds. The q^2 , R^2 , the number of compounds in the test set (n), and the number of descriptors ($nvar$)³⁸ were used as criteria for selecting the top models. Generally, a predictive model was expected to have high q^2 , R^2 and n values but with low $nvar$ values, and easily interpretable descriptors. The best model selected for the hERG prediction used eight descriptors to predict 93 compounds in the training set with a q^2 value of 0.56 and 54 compounds in the test set with an R^2 value of 0.59 (Figure 2). Y-randomization experiments did not generate any acceptable models, excluding the possibility of chance correlations (data not shown).

The analysis of our QSAR models provided insights into the relationship between the hERG channel and its blockers. The descriptors selected by our best model were: diameter, a_nO, PEOE_VSA-6, KierFlex, a_acid, PC-, SlogP and SMR_VSA6 (Supplementary Table 2A). These normalized descriptors were plotted against hERG activity. Supplementary Figure 3 shows that the hERG pIC₅₀ value increases with the diameter, KierFlex and SlogP, but decreases with the PEOE_VSA-6 and a_acid descriptors. This is in agreement with generally accepted observations that a hERG inhibitor needs to have one or two hydrophobic moieties to participate in π -stacking interactions with Phe656 in the hERG channel⁶¹. Hence, high lipophilicity (SlogP) increases the hERG liability. Moreover, basic compounds tend to be hERG active, while acidic groups mitigate hERG activity²⁹. Hence, Supplementary Figure 3C shows that most hERG active compounds do not have acidic atoms (a_acid = 0), and moreover for those compounds that contain acidic atoms, a higher number of acidic atoms (a_acid) leads to lower hERG liability. In addition, a hERG inhibitor should be long and flexible so that it can orient itself along the pore axis of the hERG channel, with its lipophilic end facing the cytosol and its polar tail facing the selectivity filter area of the hERG pore domain. Therefore, compounds with high diameter and KierFlex value have a high hERG liability. In conclusion, according to the descriptors selected in our QSAR models, a potential hERG inhibitor needs to be long and flexible, possesses hydrophobic moieties interacting with Phe656, occupies a positive charge interacting with Tyr652, and has a polar tail interacting with Thr623 and Val625. All these structural features reflect the general architecture of the hERG channel.

Dataset III contains compounds studied by Ekins et al.³², and the generated kNN models were used to predict their hERG pIC₅₀ values. We found that the correlation is poor with $R^2 = 0.21$, and overall the predicted values were much higher than the experimental values reported by Ekins et al. Some of the compounds in **Dataset III** were found to have very different pIC₅₀ values than those published by other groups^{1, 5, 11, 27–32}. After careful evaluation of these data, we found the Ekins data has much lower pIC₅₀ values^{1, 5, 11, 27–32}, in agreement with our predictions. Therefore, we think the data from other groups are more trustworthy in our case. For the eight compounds among them, if we use the data from other

groups (Supplementary Table 3), the R^2 for their predicted and corrected experimental values was increased from 0.21 to 0.66.

kNN Classification Models for Discrimination of hERG Blockers

Data Sets I and II were used to generate classification models for prediction of hERG liability. The best classification model used four descriptors. They were SMR_VSA4, radius, GCUT_SMR_0, and GCUT_SLOGP_3. The meaning of these descriptors is described in Supplementary Table 2B. It demonstrated that polarizability (SMR_VSA4, GCUT_SMR_0), lipophilicity (GCUT_SLOGP_3) and size (radius) are properties highly correlated with the hERG liability. The domperidone (**Data Set I**) and seven compounds (bismuth, bismuth subsalicylate, flucytosine, hydroxyurea, methadyl acetate, modafinil, valganciclovir) from **Data Set II** were characterized as strong outliers with significant dissimilarity to the training dataset based on our applicability domain^{62, 63}. The model selected for the hERG blocker classification had 389 compounds in the training set, including 111 actives and 278 inactives. The recall rate (sensitivity or true positive: the proportion of hERG-active inhibitors that were correctly identified) was 83.8%, as were 226 inactives (specificity = 81.3%) correctly classified. The confusion matrix of the model is shown in Supplementary Table 4A. The ability of the model to correctly predict whether a compound is hERG active or not (accuracy) is 82.0%. The model was applied to the test data set containing 101 compounds, where 35 were actives and 66 were inactives. Among those, 82.7% of the actives and 83.3% of the inactives were correctly predicted. Moreover, all hERG inhibitors with pIC_{50} higher than 5.5 were correctly classified except loratadine (Claritin). Different hERG pIC_{50} values were reported for loratadine, with 6.70 by Cavalli¹, 5.64 by Song¹¹, and 5.20 by Gavaghan²⁷. According to the Canadian FDA report⁶⁴, loratadine does not significantly block hERG potassium channels under the same *in vitro* conditions for terfenadine which showed a strong hERG inhibition⁶⁴. Furthermore, Loratadine does not have a long shape as other highly active hERG inhibitors do. Maybe this is why it was not accurately predicted by our kNN classification model. In order to exclude the possibility of chance correlations, Y-randomization experiments were executed and it did not yield any model that met our requirements (data not shown).

3D Pharmacophore Modeling

Different pharmacophore annotation schemes (Unified, PCHD, and PPCH) were applied in the pharmacophore elucidation module of MOE to search the best 3D pharmacophore models. The parameters used to select the model were: number of active molecules that match the query (coverage), the probability that the accuracy would be generated by chance alone (P), the accuracy of the query in matching the active compounds (acc1), and the accuracy of the query not matching the inactive compounds (acc0).

For the hERG blockers with basic charged moieties, the eight most active compounds with pIC_{50} values greater than 6.9 and 191 inactive compounds were employed to create the 3D pharmacophore. A model generated using the “Unified” pharmacophore scheme was characterized as the most reasonable by its high coverage (8.00), low P (−2.75), and high acc1 (1) and acc0 (0.67) values. In order to refine the pharmacophore model, features obtained in this model were compared with those obtained using the *pharmacophore consensus* method where the conformation alignment of the eight active compounds were employed to generate pharmacophore features. Subsequently, the radius of each pharmacophore feature was manually modified until it could cover as many hERG actives as possible and be able to distinguish the actives from inactives. The final query has six structural features (Figure 3A), consisting of four hydrophobic centroids, a cationic atom or hydrogen donor, and an aromatic or planar ring system. The mapping between the conformations of the eight active compounds in the training dataset and the pharmacophore

model is showed in Figure 3B. A cationic atom/hydrogen-bond donor (F4[Cat/Don]) and a hydrophobic atom (F5 [Hyd]) were features in our model that were found most important for the hERG-binding activity, in agreement with other reports^{1, 13}. Moreover, the hydrogen-bond acceptor, one of the pharmacophore features identified by Goldman¹³, has a position similar to our F3 hydrophobic centroid (F3 Hyd); this position is between the positively charged nitrogen and the aromatic ring. However, due to the differences between our active compounds in the training data set and those of Goldman's, the acceptor was not suggested in our case as a common feature by either the pharmacophore elucidation module or the consensus method. Instead, a F3 (Hyd) position was generated to improve the ability of the model to distinguish the actives from inactives. The quality of our pharmacophore models was further assessed using the *Pharmacophore Search* function for virtual screening. The confusion matrix of the model is shown in Supplementary Table 4B. Additionally, of the highly active hERG blockers with a $pIC_{50} > 6.9$ ($n = 26$), the prediction accuracy (the proportion of the total number of predictions that were correct) is 96.2% and the recall rate (sensitivity or true positive) is 92.3%. Among them, tecastemizole and n-desbutylhalofantrine ($pIC_{50} > 6.5$) were the only two compounds falsely predicted to be inactive (false negatives), probably because neither of them had an aromatic or planar ring in the F2 position and a hydrophobic center at either the F3 or F6 position (n-desbutylhalofantrine: Figure 3C; tecastemizole: Figure 3D). This low rate of false negative is what we need in hERG modeling (and other toxicity prediction) because we want to miss as less toxic compounds as possible in drug development to avoid potential toxic medications. When applied to the external **Data Set III**, 15 out of 24 hERG blockers with the amine moiety were correctly identified.

For the compounds without basic moieties, 10 active hERG blockers and 160 inactives (**Data Set VII**) were compiled as a training set for the generation of the pharmacophore model. The PCHD was identified as the best pharmacophore scheme. The model selected for prediction had high coverage (9.00), low P (−2.26), high acc1 (0.90), and high acc0 (0.66) values. The final model has seven pharmacophore features (Figure 4). Among them, the hydrogen bond acceptor and hydrophobic/aromatic features were also reported as being important for uncharged hERG blockers by Aronov et al.¹⁴ Figure 4B showed the ketone oxygen of doxazosin matches the hydrogen bond acceptor feature and its projected point, while its quinazoline, piperazine, and benzodioxine groups provide four hydrophobic features and one aromatic ring center projection point. Among the actives, 80% of the 10 compounds with pIC_{50} values greater than 5.5 matched the pharmacophore model. **Data Sets VI** and **VII** were searched using the pharmacophore query. The confusion matrix of the model is shown in Supplementary Table 4C. Although the total accuracy and specificity are reasonable (75.8% and 80.0%, respectively), the recall rate (sensitivity or true positives) is relatively low (55.9% or 19 actives out of 34). This is probably due to the small number of active compounds used for pharmacophore generation in comparison to large number of inactive compounds (10 vs. 160).

Consensus results using kNN Classification and 3D Pharmacophore Models

Based on the developed kNN classification method and 3D pharmacophore models, a consensus method was suggested for predicting the activity of the hERG blockers. For toxicity prediction, we especially prefer the maximum coverage of the real hERG inhibitors (high sensitivity) in order not to miss as less hERG-actives as possible to avoid potential toxic agents. We also prefer the negative predictive value (NPV) over 95% so that only a very small portion of toxic agents (<5%) will be forwarded to clinical use. According to this conception the hERG active compounds were those identified as positives by either kNN or pharmacophore methods as shown in Table 2A. Table 2B demonstrates that the consensus prediction using these two methods led to dramatically improved recall rate

(sensitivity=91.8%) of hERG blockers, with a slight increase of the prediction accuracy compared to pharmacophore modeling (70.3% vs. 67.5%). NPV of consensus model was augmented to 94.8%. Therefore, this consensus model is able to cover 91% of hERG blockers, whereas it achieves 95% prediction accuracy for non-hERG blockers. Unfortunately, the specificity slightly decreased due to less coverage of the hERG-inactive agents. On the other hand, if we define the hERG active compounds are those identified as positives by both kNN and pharmacophore methods (data not shown), the accuracy, sensitivity and specificity are obtained as 83.1%, 69.4% and 88.9%, respectively. Of note, with the new scheme we proposed (Supplementary Table 6), we achieved 90.2% in accuracy, 89.5% in sensitivity and 90.5% in specificity (Supplementary Table 4E) for a majority of dataset (77.6%) consensually predicted by both kNN and pharmacophore methods. It implies that if a compound receives consensus prediction by both kNN and pharmacophore modeling, we are ~90% confident to believe this prediction is accurate. Finally, our kNN QSAR method could achieve the best balance in terms of the above three parameters (Table 2B). The confusion matrix of the consensus model is shown in Supplementary Table 4D.

Homology Modeling of hERG Channel—Analogous to the Kv family of potassium channels, hERG contains a pore domain that is involved in drug binding. Electrophysiology experiments have shown that the pore domain must be in the open state for compounds to enter^{19, 65}. However, once a compound is in the cavity region, the four S6 helices can close to various degrees, depending on the bound ligand. To address the flexibility of the hERG channel, multiple homology models were generated to represent different open-close states. The binding energies for a diverse set of hERG blockers were calculated for each model, and were then validated (via PLS regression) with the experimental pIC₅₀ values of these blockers to determine which model best explains the conformation of the hERG channel in the bound form.

hERG homology models of close and open states based on the KcsA and MthK structures, along with four additional models representing the different open-close intermediate states of hERG (Supplementary Figure 5), were constructed and optimized. Compared with close state of hERG, S6 helices bend approximately 40° at Gly648 towards solvents, thus opening up a channel for ligands entry. Intermediate states are predicted states in between fully opened state and close state by bending S6 helices to different degrees. The question of whether these models can represent the hERG channel in the biological state was investigated by docking the 12 hERG blockers into the homology models and correlating their binding free energetic terms with pIC₅₀ values. The docking poses were selected according to the pharmacophore match between the ligands and those relevant residues based on the mutagenesis data in Table 1A^{19, 66, 67} of drugs blocking the hERG channel.

First, we investigate the energy values of the MacroModel calculation for the ligand and hERG complex based on the closed state (from KcsA) (Table 1B). We used SIMCA⁶⁸ to detect the outliers and established the linear correlation between the experimental and predicted pIC₅₀ values of the hERG blockers. Supplementary Figure 2A and 2B demonstrated the contribution of all descriptors to IC₅₀ using the 95% tolerance interval. PLS analysis was performed and yielded a model (Figure 5) with three principle components for the 12 compounds with $q^2 = 0.68$, $R^2 = 0.81$, and RMSE (root mean squared error) of 0.62.

$$\text{pIC}_{50} = 0.0282843 * \Delta\text{stretch} + 0.0180008 * \Delta\text{bend} + 0.00927848 * \Delta\text{torsion} - 0.00771981 * \Delta\text{VDW} + 0.00153115 * \Delta\text{electrostatic} - 61.0522 \quad \text{Equation 2}$$

Using Equation 2, the pIC_{50} values of all compounds were predicted within one log unit of the experimental value (Table 1B). Descriptors used in Equation 2 are shown in the VIP plot⁶⁰ (Supplementary Figure 2C) according to their capacity in explaining the hERG activity. Both bonded (bend, torsion and stretching) and nonbonded (electrostatic and van der Waals) interactions have VIP values greater than 0.8, which means that both types of interactions play an important role in the formation of the ligand-hERG complex. For the nonbonded term, the electrostatic interaction has a larger impact on binding than the van der Waals interaction. The larger the electrostatic interaction between the protein and ligand, the higher the hERG affinity is. This is consistent with the observation that π -stacking between the aromatic groups of the ligand with Phe656 and Tyr652 of hERG is required for strong binding^{19, 66}.

We also investigated whether the pIC_{50} values of the hERG blockers can be correctly predicted using hERG open-close intermediate state homology models for docking and binding energy calculation. Unfortunately, no acceptable PLS regression models were found; therefore, only the homology model of hERG closed form which was based on KcsV structure was used for modeling the hERG pore domain. The stereo chemical quality of the homology model was checked using PROCHECK⁶⁹. No distorted dihedral angles were detected in the Ramachandran plot (Supplementary Figure 4), and 94.4% of the residues in this homology model were in the most favored region, while 5.6% were in the allowed regions.

For the majority of these 12 compounds with known pIC_{50} , the docking poses with the highest GOLD scores matched the reported hERG mutagenesis data (Table 1A)^{19–26}. Figure 6 demonstrates that the binding mode of ibutilide which agrees with its experimental mutagenesis data (Table 1A). It has its phenyl ring as a hydrophobic group close to Tyr652 and Phe656. Tyr652 and Phe656 are two residues located in the S6 transmembrane domain of the hERG channel and have been shown in the site-directed mutagenesis^{19, 66, 67} to be critical in ligand-binding. The basic nitrogen atom of ibutilide makes a cation- π interaction with Tyr652. Its sulfonyl group acts as an acceptor and forms a hydrogen bond with the hydroxyl group of Ser624; interactions with this residue is reported to have pronounced effects on hERG blocking by ibutilide^{70, 71}. These consistencies validated our homology model, and our ability to predict the correct binding mode using the homology models is crucial for guiding the synthesis of inhibitors with less hERG liability.

Enrichment analysis and comparison of QSAR, pharmacophore, and molecular docking methods in ranking the hERG inhibitors

To rank the hERG channel blockers, three methods were employed: kNN QSAR approach, 3D pharmacophore modeling, and homology-based GOLD docking. With the kNN method, the potency of hERG inhibitors was estimated using the predicted pIC_{50} values. For the 3D pharmacophore models, the root mean squared distance between the query features and their matching ligand annotation points was used as a criterion to assess the fit of a molecule to the pharmacophore features required for hERG toxicity. With molecular docking method, we docked flexible ligand to rigid receptor based on the finding that only minor side-chain movements could occur on Phe656 and Tyr652 using flexible receptor docking by GOLD, and these local rearrangements could not affect the generation of docking poses (data not shown). The movements of these two aromatic residues indicate their critical role in ligand binding, which was consistent to the finding by Durdagi et al.¹⁸ as well as our pharmacophore modeling. GOLD fitness was used as a parameter to describe the interaction between the ligand and the hERG channel.

Enrichment³⁶ was used to evaluate the ability of these three methods in ranking the 147 hERG inhibitors out of 498 compounds (**Data Sets I and II**). The percentage of true hERG

inhibitors (Y-axis) identified using each method was plotted against their ranking. The ideal curve (black line in Figure 7) demonstrated the case where all actives can be recovered in the top 147 hits, and the gray dashed line represents a random screening. In comparison with kNN QSAR and GOLD docking, pharmacophore modeling performed slightly better for identification of top ranked true hERG inhibitors (Figure 7). Unfortunately, this function on MOE can only be used to rank those compounds predicted as hERG active as it does not give the score (RMSD) for compounds predicted as inactive (dashed orange line in Figure 7). As demonstrated, with our pharmacophore model 115 out of 147 hERG blockers (78.2%) are true positives. Moreover, the enrichment plot shows all three methods (kNN, docking, and pharmacophore) have similar performance but significantly better than random screening. Since the mechanisms of each method of modeling are different, they have their respective advantages and disadvantages: kNN machine learning methodology uses 2D molecular descriptors of compounds for its predictions; pharmacophore modeling requires their 3D conformations; whereas molecular docking needs the accurate protein structure (homology model) for the analysis of binding affinities. More importantly, the current docking/scoring are derived using true receptor-ligand binding data, and they usually need more expert knowledge of the modeling system to obtain accurate predictions. Although they have achieved great successes in “specificity-focused” studies (e.g., in virtual screening to eliminate as many true negatives as possible), we suggest that its application to “sensitivity-focused” toxicity predictions (to cover as many true positives as possible) be used with cautions. A strong hERG blocker can be easily interpreted by its interaction with hERG in a docking study; but an active compound can also be perceived as a weak binder due to various reasons such as the lack of correct receptor conformations and wrong parameterization. If this happens, then a toxic hERG blocker may go to bedside which is dangerous and is not what we want. To overcome this problem, we suggest ligand-based methods, such as the kNN and pharmacophore modeling, be used in this study. They require less expert knowledge (e.g., only 2D structures), and take into account both active and inactive hERG inhibitors. Therefore, due to the complementarities between ligand-based and structure-based methods, the combined approach would give better results.

Conclusion

We describe a comprehensive assessment of different approaches to prediction hERG liability by combining, kNN regression and classification, 3D pharmacophore modeling, homology modeling of hERG 3D structures, and molecular docking. Ligand-based approaches such as kNN and pharmacophore modeling can be employed as a predictive filter for efficient screening of large chemical databases. Structure-based model of the hERG pore domain was constructed to investigate hERG-inhibitor interactions and we could appropriately explain the binding of different inhibitors as well as the reported mutagenesis studies. Therefore we are convinced that reliable structural models have been obtained for hERG open and closed forms. In summary, these *in silico* models of hERG and its blockers are useful and can be employed to efficiently screen hERG liabilities in early lead discovery as well as to guide medicinal synthesis for reduce hERG binding during lead optimization stage.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the US Department of Defense Concept Awards (BC085871), Grant # IRG-08-061-01 from the American Cancer Society and MDACC-UT Austin CTT-TI3D grants. We thank OpenEye

Scientific and ChemAxon for providing free academic licenses of their software packages. We also thank Dr. Alexander Tropsha for letting us use kNN program and Mr. John Morrow for proof reading of the manuscript.

Abbreviations

hERG	the human <i>Ether-à-go-go</i> related gene
QSAR	quantitative structure-activity relationship
kNN	k nearest neighbor
PCA	principle component analysis
ANN	artificial neural network
SVM	support vector machine
PLS	partial least squares
MD	molecular dynamics
PPV	positive predictive value
NPV	negative predictive value

References

- De Ponti F, Poluzzi E, Cavalli A, Recanatini M, Montanaro N. Safety of non-antiarrhythmic drugs that prolong the QT interval or induce torsade de pointes: An overview. *Drug Safety*. 2002; 25(4): 263–286. [PubMed: 11994029]
- Pearlstein R, Vaz R, Rampe D. Understanding the structure-activity relationship of the human ether-a-go-go-related gene cardiac K⁺ channel. A model for bad behavior. *J Med Chem*. 2003; 46(11): 2017–2022. [PubMed: 12747773]
- Brown AM. Drugs, hERG, and sudden death. *Cell Calcium*. 2004; 35(6):543–547. [PubMed: 15110144]
- Netzer R, Ebneith A, Bischoff U, Pongs O. Screening lead compounds for QT interval prolongation. *Drug Discov Today*. 2001; 6(2):78–84. [PubMed: 11166255]
- Yoshida K, Niwa T. Quantitative structure-activity relationship studies on inhibition of hERG potassium channels. *J Chem Inf Model*. 2006; 46(3):1371–8. [PubMed: 16711756]
- Roche O, Trube G, Zuegge J, Pflimlin P, Alanine A, Schneider G. A virtual screening method for prediction of the hERG potassium channel liability of compound libraries. *ChemBioChem*. 2002; 3(5):455–459. [PubMed: 12007180]
- Shen MY, Su BH, Esposito EX, Hopfinger AJ, Tseng YJ. A comprehensive support vector machine binary hERG classification model based on extensive but biased end point hERG data sets. *Chem Res Toxicol*. 2011; 24(6):934–49. [PubMed: 21504223]
- Su BH, Shen MY, Esposito EX, Hopfinger AJ, Tseng YJ. In silico binary classification QSAR models based on 4D-fingerprints and MOE descriptors for prediction of hERG blockage. *J Chem Inf Model*. 2010; 50(7):1304–18. [PubMed: 20565102]
- Ekins S, Chekmarev DS, Kholodovych V, Balakin KV, Ivanenkov Y, Welsh WJ. Shape signatures: New descriptors for predicting cardiotoxicity in silico. *Chemical Research in Toxicology*. 2008; 21(6):1304–1314. [PubMed: 18461975]
- Keseru GM. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg Med Chem Lett*. 2003; 13(16):2773–5. [PubMed: 12873512]
- Song M, Clark M. Development and Evaluation of an in Silico Model for hERG Binding. *J Chem Inf Model*. 2006; 46(1):392–400. [PubMed: 16426073]
- Ekins S, Crumb WJ, Sarazan RD, Wikel JH, Wrighton SA. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go-related gene potassium channel. *Journal of Pharmacology and Experimental Therapeutics*. 2002; 301(2):427–434. [PubMed: 11961040]

13. Aronov AM, Goldman BB. A model for identifying HERG K⁺ channel blockers. *Bioorg Med Chem*. 2004; 12(9):2307–15. [PubMed: 15080928]
14. Aronov AM. Common pharmacophores for uncharged human ether-a-go-go-related gene (hERG) blockers. *J Med Chem*. 2006; 49(23):6917–21. [PubMed: 17154521]
15. Stansfeld PJ, Gedeck P, Gosling M, Cox B, Mitcheson JS, Sutcliffe MJ. Drug block of the hERG potassium channel: insight from modeling. *Proteins: Struct, Funct, Bioinf*. 2007; 68(2):568–580.
16. Rajamani R, Tounge BA, Li J, Reynolds CH. A two-state homology model of the hERG K⁺ channel: application to ligand binding. *Bioorg Med Chem Lett*. 2005; 15(6):1737–1741. [PubMed: 15745831]
17. Taboureau O, Li QY, Jorgensen FS, Oprea T, Brunak S. hERG classification model based on a combination of support vector machine method and GRIND descriptors. *Molecular Pharmaceutics*. 2008; 5(1):117–127. [PubMed: 18197627]
18. Durdagi S, Duff HJ, Noskov SY. Combined receptor and ligand-based approach to the universal pharmacophore model development for studies of drug blockade to the hERG1 pore domain. *J Chem Inf Model*. 2011; 51(2):463–74. [PubMed: 21241063]
19. Mitcheson JS, Chen J, Lin M, Culberson C, Sanguinetti MC. A structural basis for drug-induced long QT syndrome. *Proceedings of the National Academy of Sciences of the United States of America*. 2000; 97(22):12329–12333. [PubMed: 11005845]
20. Kamiya K, Niwa R, Mitcheson JS, Sanguinetti MC. Molecular determinants of HERG channel block. *Mol Pharmacol*. 2006; 69(5):1709–16. [PubMed: 16474003]
21. Ridley JM, Milnes JT, Witchel HJ, Hancox JC. High affinity HERG K(+) channel blockade by the antiarrhythmic agent dronedarone: resistance to mutations of the S6 residues Y652 and F656. *Biochem Biophys Res Commun*. 2004; 325(3):883–91. [PubMed: 15541373]
22. Milnes JT, Crociani O, Arcangeli A, Hancox JC, Witchel HJ. Blockade of HERG potassium currents by fluvoxamine: incomplete attenuation by S6 mutations at F656 or Y652. *Br J Pharmacol*. 2003; 139(5):887–98. [PubMed: 12839862]
23. Witchel HJ, Dempsey CE, Sessions RB, Perry M, Milnes JT, Hancox JC, Mitcheson JS. The low-potency, voltage-dependent HERG blocker propafenone--molecular determinants and drug trapping. *Mol Pharmacol*. 2004; 66(5):1201–12. [PubMed: 15308760]
24. Kamiya K, Mitcheson JS, Yasui K, Kodama I, Sanguinetti MC. Open channel block of HERG K(+) channels by vesnarinone. *Mol Pharmacol*. 2001; 60(2):244–53. [PubMed: 11455010]
25. Lees-Miller JP, Duan Y, Teng GQ, Duff HJ. Molecular determinant of high-affinity dofetilide binding to HERG1 expressed in *Xenopus* oocytes: involvement of S6 sites. *Mol Pharmacol*. 2000; 57(2):367–74. [PubMed: 10648647]
26. Sanchez-Chapula JA, Ferrer T, Navarro-Polanco RA, Sanguinetti MC. Voltage-dependent profile of human ether-a-go-go-related gene channel block is influenced by a single residue in the S6 transmembrane domain. *Mol Pharmacol*. 2003; 63(5):1051–8. [PubMed: 12695533]
27. Gavaghan CL, Arnby CH, Blomberg N, Strandlund G, Boyer S. Development, interpretation and temporal evaluation of a global QSAR of hERG electrophysiology screening data. *J Comput-Aided Mol Des*. 2007; 21(4):189–206. [PubMed: 17384921]
28. Keseru GM. Prediction of hERG potassium channel affinity by traditional and hologram qSAR methods. *Bioorg Med Chem Lett*. 2003; 13(16):2773–2775. [PubMed: 12873512]
29. Sun H. An accurate and interpretable Bayesian classification model for prediction of hERG liability. *ChemMedChem*. 2006; 1(3):315–322. [PubMed: 16892366]
30. Recanatini M, Poluzzi E, Masetti M, Cavalli A, De Ponti F. QT prolongation through hERG K(+) channel blockade: current knowledge and strategies for the early prediction during drug development. *Med Res Rev*. 2005; 25(2):133–66. [PubMed: 15389727]
31. Pearlstein RA, Vaz RJ, Kang J, Chen X-L, Preobrazhenskaya M, Shchekotikhin AE, Korolev AM, Lysenkova LN, Miroshnikova OV, Hendrix J, Rampe D. Characterization of HERG potassium channel inhibition using CoMSiA 3D QSAR and homology modeling approaches. *Bioorg Med Chem Lett*. 2003; 13(10):1829–1835. [PubMed: 12729675]
32. Ekins S, Balakin KV, Savchuk N, Ivanenkov Y. Insights for Human Ether-a-Go-Go-Related Gene Potassium Channel Inhibition Using Recursive Partitioning and Kohonen and Sammon Mapping Techniques. *J Med Chem*. 2006; 49(17):5059–5071. [PubMed: 16913696]

33. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* 2008; 36(Database Iss):D901–D906. [PubMed: 18048412]
34. Morgan TK Jr, Sullivan ME. An overview of class III electrophysiological agents: a new generation of antiarrhythmic therapy. *Prog Med Chem.* 1992; 29:65–108. [PubMed: 1475373]
35. Golbraikh A, Shen M, Xiao Z, Xiao Y-D, Lee K-H, Tropsha A. Rational selection of training and test sets for the development of validated QSAR models. *J Comput-Aided Mol Des.* 2003; 17(2–4):241–253. [PubMed: 13677490]
36. Du-Cuny L, Song Z, Moses S, Powis G, Mash EA, Meuliet EJ, Zhang S. Computational modeling of novel inhibitors targeting the Akt pleckstrin homology domain. *Bioorg Med Chem.* 2009; 17(19):6983–6992. [PubMed: 19734051]
37. Sharaf MA, Illman DL, Kowalski BR. Chemical Analysis. *Chemometrics.* 1986; 82:332.
38. Zhang S, Golbraikh A, Oloff S, Kohn H, Tropsha A. A Novel Automated Lazy Learning QSAR (ALL-QSAR) Approach: Method Development, Applications, and Virtual Screening of Chemical Databases Using Validated ALL-QSAR Models. *J Chem Inf Model.* 2006; 46(5):1984–1995. [PubMed: 16995729]
39. Zheng W, Tropsha A. Novel variable selection quantitative structure-property relationship approach based on the K-nearest-neighbor principle. *J Chem Inf Comput Sci.* 2000; 40(1):185–194. [PubMed: 10661566]
40. Testai L, Bianucci AM, Massarelli I, Breschi MC, Martinotti E, Calderone V. Torsadogenic cardiotoxicity of antipsychotic drugs: A structural feature, potentially involved in the interaction with cardiac HERG potassium channels. *Curr Med Chem.* 2004; 11(20):2691–2706. [PubMed: 15544470]
41. Ok D, Li C, Abbadie C, Felix JP, Fisher MH, Garcia ML, Kaczorowski GJ, Lyons KA, Martin WJ, Priest BT, Smith MM, Williams BS, Wyvrat MJ, Parsons WH. Synthesis and SAR of 1,2-trans-(1-hydroxy-3-phenylprop-1-yl)cyclopentane carboxamide derivatives, a new class of sodium channel blockers. *Bioorg Med Chem Lett.* 2006; 16(5):1358–1361. [PubMed: 16337121]
42. Sisko JT, Tucker TJ, Bilodeau MT, Buser CA, Cieccko PA, Coll KE, Fernandes C, Gibbs JB, Koester TJ, Kohl N, Lynch JJ, Mao X, McLoughlin D, Miller-Stein CM, Rodman LD, Rickert KW, Sepp-Lorenzino L, Shipman JM, Thomas KA, Wong BK, Hartman GD. Potent 2-[(pyrimidin-4-yl)amine]-1,3-thiazole-5-carbonitrile-based inhibitors of VEGFR-2 (KDR) kinase. *Bioorg Med Chem Lett.* 2006; 16(5):1146–1150. [PubMed: 16368234]
43. Friesen RW, Ducharme Y, Ball RG, Blouin M, Boulet L, Cote B, Frenette R, Girard M, Guay D, Huang Z, Jones TR, Laliberte F, Lynch JJ, Mancini J, Martins E, Masson P, Muise E, Pon DJ, Siegl PKS, Styhler A, Tsou NN, Turner MJ, Young RN, Girard Y. Optimization of a Tertiary Alcohol Series of Phosphodiesterase-4 (PDE4) Inhibitors: Structure-Activity Relationship Related to PDE4 Inhibition and Human Ether-a-go-go Related Gene Potassium Channel Binding Affinity. *J Med Chem.* 2003; 46(12):2413–2426. [PubMed: 12773045]
44. Labute P, Williams C, Feher M, Sourial E, Schmidt JM. Flexible Alignment of Small Molecules. *J Med Chem.* 2001; 44(10):1483–1490. [PubMed: 11334559]
45. Lauri G, Bartlett PA. CAVEAT: a program to facilitate the design of organic molecules. *J Comput-Aided Mol Des.* 1994; 8(1):51–66. [PubMed: 8035213]
46. Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science.* 1998; 280(5360):69–77. [PubMed: 9525859]
47. Jiang Y, Lee A, Chen J, Cadene M, Chait Brian T, MacKinnon R. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature.* 2002; 417(6888):515–22. [PubMed: 12037559]
48. Pearson, WR. *ALIGN*, 2.0u66. The University of Virginia; 1997.
49. Jiang Y, Lee A, Chen J, Cadene M, Chait BT, MacKinnon R. The open pore conformation of potassium channels. *Nature (London, U K).* 2002; 417(6888):523–526. [PubMed: 12037560]
50. MODELLER [9v6]. UCSF; San Francisco, CA, USA: 2009.
51. Bower MJ, Cohen FE, Dunbrack RL Jr. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J Mol Biol.* 1997; 267(5):1268–1282. [PubMed: 9150411]

52. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr, Sect D: Biol Crystallogr*. 2004; D60(12, Pt. 1):2126–2132. [PubMed: 15572765]
53. Case David A, Cheatham Thomas E 3rd, Darden T, Gohlke H, Luo R, Merz Kenneth M Jr, Onufriev A, Simmerling C, Wang B, Woods Robert J. The Amber biomolecular simulation programs. *J Comput Chem*. 2005; 26(16):1668–88. [PubMed: 16200636]
54. Wang JM, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*. 2000; 21(12):1049–1074.
55. Jin L, Wu Y. Molecular mechanism of the sea anemone toxin ShK recognizing the Kv1.3 channel explored by docking and molecular dynamic simulations. *J Chem Inf Model*. 2007; 47(5):1967–72. [PubMed: 17718553]
56. GOLD [32]. CCDC; Cambridge, UK: 2007.
57. Aronov AM. Predictive in silico modeling for hERG channel blockers. *Drug Discov Today*. 2005; 10(2):149–55. [PubMed: 15718164]
58. Yang H, Shen Y, Chen J, Jiang Q, Leng Y, Shen J. Structure-based virtual screening for identification of novel 11beta-HSD1 inhibitors. *Eur J Med Chem*. 2009; 44(3):1167–71. [PubMed: 18653260]
59. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, Farid R, Felts AK, Halgren TA, Mainz DT, Maple JR, Murphy R, Philipp DM, Repasky MP, Zhang LY, Berne BJ, Friesner RA, Gallicchio E, Levy RM. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem*. 2005; 26(16):1752–80. [PubMed: 16211539]
60. Eriksson, L.; Johansson, N.; Kettaneh-Wold, N.; Wold, S. Multi- and Megavariate Data Analysis: Principles and Applications. Umetrics Academy; Umea: 1999.
61. Aronov AM. Predictive in silico modeling for hERG channel blockers. *Drug Discovery Today*. 2005; 10(2):149–155. [PubMed: 15718164]
62. Koziowski AP, Sun H, Brognard J, Dennis PA. Novel PI Analogues Selectively Block Activation of the Pro-survival Serine/Threonine Kinase Akt. *J Am Chem Soc*. 2003; 125(5):1144–1145. [PubMed: 12553797]
63. Friesner Richard A, Banks Jay L, Murphy Robert B, Halgren Thomas A, Klicic Jasna J, Mainz Daniel T, Repasky Matthew P, Knoll Eric H, Shelley M, Perry Jason K, Shaw David E, Francis P, Shenkin Peter S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*. 2004; 47(7):1739–49. [PubMed: 15027865]
64. Lapenta, JR. The loratadine and their 110 adverse effects. <http://www.dermagicjournal.20m.com/year2001/11B.html>
65. Spector PS, Curran ME, Keating MT, Sanguinetti MC. Class III antiarrhythmic drugs block HERG, a human cardiac delayed rectifier K⁺ channel. Open-channel block by methanesulfonanilides. *Circulation research*. 1996; 78(3):499–503. [PubMed: 8593709]
66. Lees-Miller JP, Duan Y, Teng GQ, Thorstad K, Duff HJ. Novel gain-of-function mechanism in K(+) channel-related long-QT syndrome: altered gating and selectivity in the HERG1 N629D mutant. *Circulation research*. 2000; 86(5):507–13. [PubMed: 10720411]
67. Chen J, Seeböhm G, Sanguinetti Michael C. Position of aromatic residues in the S6 domain, not inactivation, dictates cisapride sensitivity of HERG and eag potassium channels. *Proc Natl Acad Sci U S A*. 2002; 99(19):12461–6. [PubMed: 12209010]
68. SIMCA. Umetrics AB; Umea, Sweden: 2011.
69. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993; 26(2):283–91.
70. Perry M, De Groot MJ, Helliwell R, Leishman D, Tristani-Firouzi M, Sanguinetti MC, Mitcheson J. Structural determinants of HERG channel block by clofilium and ibutilide. *Mol Pharmacol*. 2004; 66(2):240–249. [PubMed: 15266014]
71. Perry M, Stansfeld PJ, Leaney J, Wood C, de Groot MJ, Leishman D, Sutcliffe MJ, Mitcheson JS. Drug binding interactions in the inner cavity of hERG channels: molecular insights from structure-activity relationships of clofilium and ibutilide analogs. *Mol Pharmacol*. 2006; 69(2):509–519. [PubMed: 16291873]

Figure 1. The hERG data sets used for the QSAR, pharmacophore and docking analysis. Data Set I

contains 147 hERG active compounds; **Data Set II** contains 351 inactive compounds; **Data Set III** contains the 31 actives compiled by Ekins et al.³²; Split from **Data Set I**, **Data Set IV** contains 113 actives with basic moieties and **Data Set VI** contains 34 actives without basic moieties; Split from **Data Set II**, **Data Set V** contains 191 inactives with basic moieties and **Data Set VII** contains 160 inactives without basic moieties.

Figure 2. Experimental versus predicted hERG pIC₅₀ using one of our QSAR models
This model has 93 compounds in the training set and 54 in the test set.

Figure 3. A. 3D pharmacophore model for compounds with basic moieties

Hyd: Hydrophobic centroid; Cat: Cationic atom; Don: H-bond donor; PiN: Aromatic or planar ring system. **B.** Mapping of the eight active compounds in the training dataset onto the 3D pharmacophore model. **C.** Structure of n-desbutylhalofantrine fitted to the pharmacophore model. **D.** Structure of tecastemizole fitted to the pharmacophore model.

Figure 4. A. 3D pharmacophore model for compounds without basic moieties
Acc: H-bond acceptor; Acc2: H-bond acceptor projected point; Aro: Aromatic ring center;
Aro2: Aromatic ring center projected point; Hyd: Hydrophobic region. **B.** Conformation of
doxazosin mapped onto our seven-point pharmacophore model.

Figure 5.
Prediction of pIC₅₀ for 12 compounds with docking poses selected based on the mutagenesis data.

Figure 6. The docked pose of ibutilide (in magenta sticks) in the cavity of the hERG channel (green ribbons)

It interacts strongly with residues Tyr652, Phe656 and Ser624 (displayed in line), which is in agreement with the mutagenesis data.

Figure 7. The enrichment plot of hERG inhibitors obtained from kNN QSAR, 3D pharmacophore, and GOLD docking methods

The gray dashed line represents random screening. The X-axis shows the compound ranking. The Y-axis shows the percentage of true actives predicted. For pharmacophore modeling, the program (MOE) only gave ranking for the top 248 hits which were identified as hERG-active. So the rest were represented with the orange dashed line for completeness.

Table 1A

Experimental mutagenesis and pIC₅₀ data of 12 hERG blockers.

Compound	Structure	pIC ₅₀ (Exp)	Residues
e4031		7.74	T623, S624, V625, G648, Y652, F656, V659 ²⁰
ibutilide		7.82	T623, S624, V625, G648, Y652, F656, V659 ²⁰
dronedarone		4.23	F656 ²¹
mk499		7.47	T623, S624, V625, G648, Y652, F656, V659 ²⁰

Compound	Structure	pIC ₅₀ (Exp)	Residues
amiodarone		5.00	Y652, F656 ²¹
fluvoxamine		4.80	S631 ²²
propafenone		6.36	F656 ²³
vesnarinone		5.96	G628, S631 ²⁴

Compound	Structure	pIC ₅₀ (Exp)	Residues
terfenadine		6.70	Y652, F656 ¹⁹
cisapride		7.35	Y652, F656 ¹⁹
quinidine		6.49	F656 but not S631 ²⁵
chloroquine		5.60	Y652, F656 ²⁶

Table 1B
Calculated energetic terms and predicted pIC_{50} for 12 compounds docked into the hERG homology model.

Compound	$\Delta stretch(kJ/mol)$	$\Delta bend(kJ/mol)$	$\Delta torsion(kJ/mol)$	$\Delta improper\ torsion(kJ/mol)$	$\Delta VDW(kJ/mol)$	$\Delta electrostatic(kJ/mol)$	$pIC_{50}\ (Pred)$	$ApIC_{50}(Exp-Pred)$
e4031	1963.11	450.26	1217.21	48.38	-1285.49	-11115.75	6.78	0.96
ibutilide	1967.33	439.19	1253.39	49.69	-1360.04	-11500.08	7.02	0.80
dronedarone	1872.40	486.90	1160.45	44.68	-1430.63	-11979.89	4.14	0.09
mk499	1924.13	494.60	1194.96	43.61	-1418.60	-11282.03	7.04	0.43
amiodarone	1931.13	443.83	1178.21	43.28	-1313.19	-11378.87	5.20	-0.20
fluvoxamine	1978.35	473.81	1023.87	51.95	-1188.80	-11393.08	4.67	0.13
propafenone	2002.00	393.07	1044.21	39.84	-1224.08	-9887.91	6.65	-0.29
vesnarinone	1919.92	474.72	1301.54	47.78	-1401.05	-11608.96	6.91	-0.95
terfenadine	1939.98	489.88	1109.64	45.58	-1428.26	-11126.23	6.92	-0.22
cisapride	1915.36	521.30	1358.98	57.91	-1449.23	-12169.97	7.67	-0.32
quinidine	2058.25	388.36	968.05	36.35	-1072.18	-9693.92	6.57	-0.08
chloroquine	2011.19	400.62	1120.64	47.96	-1168.41	-10785.15	5.95	-0.35

Note: $\Delta stretch$, $\Delta bend$, $\Delta torsion$, $\Delta improper\ torsion$, ΔVDW , $\Delta electrostatic$ are estimated interaction energy difference for each ligand in complex and free state. Mutagenesis data were reported for individual drug on their ability to block hERG functions.

Table 2A

The consensus scheme of classifying hERG-actives and -inactives using kNN QSAR and 3D pharmacophore modeling methods.

kNN	Pharmacophore	Consensus
1	1	1
0	1	1
1	0	1
0	0	0

Note: “1” indicates predicted as hERG-positive and “0” represents predicted as hERG-negative.

Comparison of kNN QSAR, 3D pharmacophore, and consensus models in classification of hERG-actives and inactives.

Table 2B

Method	hERG Activity	N _{Exp}	N _{Pred}	Accuracy	Sensitivity	Specificity	NPV
kNN classification	Active	147	122	82.2%	83.0%	80.1%	91.8%
	Inactive	351	281				
Pharmacophore model	Active	147	115	67.5%	78.2%	63.0%	87.4%
	Inactive	351	221				
Consensus method	Active	147	135	70.3%	91.8%	62.7%	94.8%
	Inactive	351	220				

Note: NPV : negative predictive value; N_{Exp} is the number of compounds experimentally observed. N_{Pred} is the number of active or inactive compounds correctly predicted.