CHEMMEDCHEM

DOI: 10.1002/cmdc.200700221

A Composite Model for hERG Blockade

Christian Kramer,[a, b] Bernd Beck,*[a] Jan M. Kriegl,[a] and Timothy Clark*[b]

Dedicated to the memory of Professor Martyn Ford

hERG blockade is one of the major toxicological problems in lead structure optimization. Reliable ligand-based in silico models for predicting hERG blockade therefore have considerable potential for saving time and money, as patch-clamp measurements are very expensive and no crystal structures of the hERG-encoded channel are available. Herein we present a predictive QSAR model for hERG blockade that differentiates between specific and nonspecific binding. Specific binders are identified by preliminary pharmacophore scanning. In addition to descriptor-based models for the compounds selected as hitting one of two different pharmacophores, we also use a model for nonspecific binding that reproduces blocking properties of molecules that do not fit either of the two pharmacophores. PLS and SVR models based on interpretable quantum mechanically derived descriptors on a literature dataset of 113 molecules reach overall R^2 *values between 0.60 and 0.70 for independent validation sets and* R^2 *values between 0.39 and 0.76 after partitioning according to the pharmacophore search for the test sets. Our findings suggest that hERG blockade may occur through different types of binding, so that several different models may be necessary to assess hERG toxicity.*

Introduction

Cardiac arrhythmia due to blockade of potassium channels coded by the human *ether-a-go-go*-related gene (hERG)[1] is a major concern for both the pharmaceutical industry and health regulatory agencies.[2–4] Since 1996, when terfenadine,[5,6] and later astemizole[7] and cisapride[8] were withdrawn from the market, attention has been focused on optimizing compounds in order to avoid prolonging the QT interval.[9] The QT interval is defined as the time interval between the start of the Q wave and the end of the T wave in the heart's electric cycle. As many drugs from different classes have been shown to prolong the QT interval because of hERG blockade,[10] hERG is considered to be a general anti-target.[11] It is now used as a surrogate marker for cardiotoxicity. In lead structure optimization, hERG blockade is tested early and monitored throughout the drug-development process.[12]

The hERG gene and its product were discovered when the genome of patients with inherited long QT syndrome (LQTS) was analyzed. Patients with LQTS exhibit a significant predisposition for cardiac arrhythmia of the *torsades des pointes* type.[13] QT prolongation can in most cases be traced back to a mutation in the hERG gene[14] that codes for the potassium channel relevant for I_{kr}, the rapid delayed rectifier current.[15,16] Recently, some other genes have been found for which mutations may lead to QT prolongation. Among them are *KCNQ1*, the gene that encodes the potassium channel relevant for I_{ks}, a smaller depolarizing K^+ current, and *SCN5A*, the gene that encodes the cardiac Na^+ channel.[17,18] However, all clinically relevant cases of drug-induced LQTS could be traced back to either hERG blockade[19,20] or interference with hERG trafficking to the cell surface.[21]

Cardiac polarization/repolarization is managed by different ion channels. In brief, polarization is caused by sodium ions entering the cell through specific sodium channels. This causes a shift of the cell wall potential towards positive values. As soon as a certain polarity has been reached, potassium channels start to open up. Positively charged potassium ions flow out of the cell, and depolarization is re-established. Neighboring voltage-dependent sodium channels also open up upon polarization, and the whole excitation process moves along the surface. There are more types of ion channels involved in the regulation of the excitation process, but the amount of charge that they transfer is comparatively small.

Most of the charge transferred in the repolarization phase is transferred via the potassium channel coded for by the hERG gene. Upon blockade, the action potential will rest longer, which results in an increased duration of the relative QT interval that can be observed in electrocardiograph (ECG) traces.

Disturbing the QT interval may lead to instability in the heart rhythm because all the electric potentials in the complex concert of the heartbeat are very finely adjusted. *Torsades de pointes*, a disordered uncontrolled excitation of cells in the

[a] C. Kramer, Dr. B. Beck, Dr. J. M. Kriegl
Department of Lead Discovery
Boehringer–Ingelheim Pharma GmbH & Co. KG, 88397 Biberach (Germany)
Fax: (+49) 7351-838-151
E-mail: bernd.beck@boehringer-ingelheim.com

[b] C. Kramer, Prof. Dr. T. Clark
Computer-Chemie-Centrum and
Interdisciplinary Center for Molecular Materials
Friedrich-Alexander Universität Erlangen–Nürnberg
Nägelsbachstrasse 25, 91052 Erlangen (Germany)
Fax: (+49) 9131-852-6565
E-mail: tim.clark@chemie.uni-erlangen.de

Supporting information for this article is available on the WWW under http://www.chemmedchem.org or from the author.

heart muscle, may occur. This cacophony of excitations prevents the heart from functioning properly and may lead to lethal ventricular fibrillation.

Biochemical details

The protein channel responsible for I_{kr} is a tetramer that consists of four identical subunits coded for by the hERG gene. As is the case for most transmembrane proteins, the hERG channel has so far resisted attempts to obtain an X-ray or NMR structure. Several homology models[9,22,23] based on the crystal structure of potassium channels from bacterial cell walls[24–28] have been published.

Important structural features

The homology models reported are all in agreement that a large cavity at the inner end of the pore with an estimated diameter of up to 12 Å, depending on the conformation, is an important feature of the hERG channel.[23] This pore is able to accommodate a wide range of structurally diverse compounds.

Mutagenesis and alanine-scanning studies have revealed the important roles of several amino acids.[29–34] Phe656 and Tyr652 have been identified as very important binding partners for almost all compounds tested[29] apart from metoprolol.[35] Additionally, depending on the compound, one or more of the amino acids Ser624, Thr623, Val625, Gly648, and Val650 are necessary for binding. In all homology models and docking studies published,[22,23,36–38] the side chains of Phe656, Tyr652, Ser624, Thr623, and Val625 point towards the inside of the pore. A schematic view of the hERG channel is shown in Figure 1.

Because of its nature as a switchable gating channel, the hERG channel can be present in various forms: open, closed, inactivated, and probably some additional intermediate states. The promiscuous nature of the hERG channel suggests that there must be many different binding motifs and geometries, each suited to individual ligands or groups of ligands.

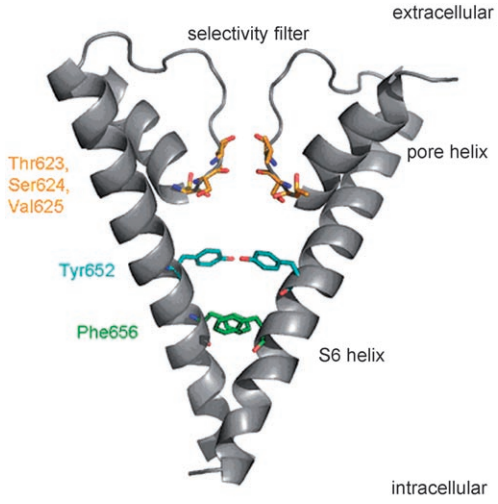

Figure 1. Schematic representation of the hERG channel, displaying only two of the four subunits. For simplicity, only the selectivity filter and the S6 and pore helices are shown. Important residues are shown in stick representation. The extracellular side is indicated at the top. The picture was generated from a homology model of the closed hERG channel (template: KcsA, PDB entry: 1K4C).

Published models

Because hERG is a very important anti-target that has caused withdrawal of several high-selling drugs from the market, many groups have published theoretical studies on hERG blockade.

Descriptor-based QSAR

Since 2002, when the first ligand-based hERG blockade QSAR study was published by Roche et al.,[39] much attention has been paid to establishing QSAR models for hERG blockade. Many public and commercially available descriptor sets have been used to model hERG IC_{50} values. Models based on 2D descriptors provided by MOE[40] and TSAR,[41] Ghose–Crippen, Kier & Hall topological indices, Isis-Keys, atom pair, electrotopological state descriptors,[42] molecular fingerprints, and molecular fragments have been published.[43–46] A large variety of statistical methods has also been used; self-organizing maps (SOMs),[39] multiple linear regression (MLR), partial least squares (PLS), logistic regression,[47] support vector machines (SVMs), Bayesian classifiers, and decision-tree algorithms. However, the resulting models can seldom be compared directly because they were built using different datasets. Usually, models based on proprietary unpublished datasets have better R^2 values than those based on publicly available datasets.[44] The benchmark of serious published-data-based QSAR studies that use nonlinear relationships is around $R^2 = 0.70$.[48]

Pharmacophore models and 3D QSAR

Cavalli et al.[49] were the first to publish a hERG pharmacophore that resulted from a comparative molecular field analysis (CoMFA) study in 2002. Since then, further pharmacophore studies have been published, including comparative molecular similarity indices (CoMSIA)[9] and other 3D QSAR[50] studies on hERG activity prediction.[51] All pharmacophores published so far include some common features: a positively ionizable nitrogen feature in the center and two or three hydrophobic–aromatic or hydrophobic features on the periphery, distributed tetragonally around the nitrogen. Aronov recently published a pharmacophore for hERG blockers based on proprietary data that does not contain the positively ionizable nitrogen.[52] The most active molecule known in the literature without the nitrogen feature is ketoconazole, with $IC_{50} = 1.9$ μM. Leong recently published a pharmacophore ensemble/SVM model for hERG inhibition with R^2 values > 0.90 for a test set of 13 molecules.[53]

Herein we present a regression study of modeling hERG blockade. We have used quantum mechanically derived values of local properties on the molecular surface, shape and charge autocorrelation descriptors, and some basic 1D and 2D descriptors. The PLS and ε-support vector regression (ε-SVR) algo-

rithms were used to build the regression models. These techniques facilitate the interpretation of important descriptors and provide highly predictive models for the test sets.

We first describe novel descriptors that are relevant for predicting hERG blockade. However, we also describe a method of combining pharmacophores and traditional descriptor-based QSAR that subdivides compounds into specific and unspecific blockers. Composite regression models reach R^2y values for the literature dataset that are equal to R^2 values of previously published PLS models. Further composite models give insight into patterns that are relevant for hERG binding for different subgroups of compounds available from the literature.

Results and Discussion

For all the compounds collected from the literature and purified for contingency and comparability, we calculated a set of descriptors based on quantum mechanical properties. They represent statistical variables that describe the distribution of the local molecular electrostatic potential (MEP), the local ionization energy (IEL), the local electron affinity (EAL), the local polarizability (POL), as well as external and internal Shannon entropies (SHANE, SHANI). The local properties projected onto the surface of cisapride are shown in Figure 2. Furthermore,

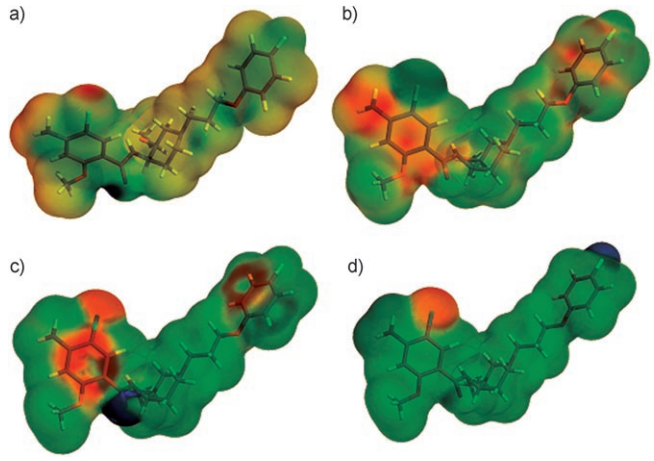

Figure 2. Projection of local properties to the surface of cisapride: a) MEP, b) IEL, c) EAL, d) POL; blue = low values, green = mean values, red = high values.

shape and charge autocorrelation descriptors and some other basic descriptors were calculated (see the Experimental Section below for a complete list and data preparation). Based on these descriptors we generated regression models.

Model 1

The best averaged R^2y value for the test sets after 20-fold random cross-validation was obtained using the ε-SVR algorithm, a polynomial kernel of degree 2, and multiple linear regression (MLR)-based descriptor selection. A validation set that was partitioned off before training is predicted with R^2y, root-mean-squared error (RMSE) and mean-unsigned error (MUE) values similar to the cross-validation results. A plot of measured vs. predicted pIC_{50} values including standard deviations is shown in Figure 3. MLR descriptor selection yielded seven descriptors: EALmax, EALmin, POLmin, SHANIbar, Naryl, shapeQ2 and shapeQ4.

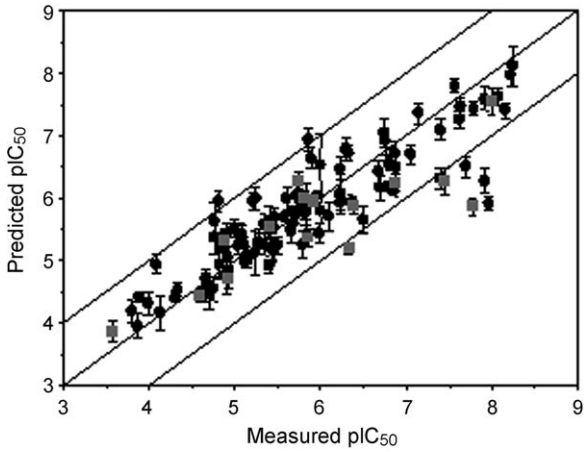

Figure 3. Mean measured and predicted pIC_{50} values for the training, test, and validation set for the MLR descriptor set including standard deviation (ε-SVR model, polynomial kernel of degree 2, split according to pIC_{50} bins); ● = test and training set, ■ = validation set. Mean R^2y(train) = 0.81, N = 75; mean R^2y(test) = 0.68, N = 23; mean RMSE train&test = 0.50; mean MUE train&test = 0.37, N = 98; mean R^2y(validation) = 0.70; mean RMSE validation = 0.74; mean MUE validation = 0.55, N = 15.

ε-SVR models are more accurate than PLS models, as measured by R^2, because nonlinear relationships between the descriptors are included. Model 1 shows that it is possible to achieve statistical values similar to the best literature-data-based models published so far with the descriptors used in this study. However, because it is easier to interpret a PLS model than an ε-SVR model, we focus our attention on PLS models for the analysis of important descriptors.

Model 2: most significant descriptors, conservative descriptor selection

In order to analyze the significance of different descriptors, the inclusion criterion for descriptors in the stepwise MLR procedure, the F value, was changed. It was defined in such a way that only relevant descriptors are included following an approach suggested by Martyn Ford and inspired by Livingstone and Salt.[54] The procedure was repeated for each of the 20 training sets. Shape, EALmin, and Naryl were always selected as descriptors under this stricter inclusion criterion. Other descriptors were only included once or never. This suggests that only shape, Naryl, and EALmin are really significant. Mean R^2y values for the 20 models of all test/training set splits built with shape, Naryl, and EALmin are R^2y(train) = 0.59 and R^2y(test) = 0.57. In adding squares and cubes of the three descriptors and executing the same procedure, only shape^2 and shape^3 are always selected. We built PLS models for all 20 test/training set

splits. Measured and predicted pIC_{50} values are plotted in Figure 4.

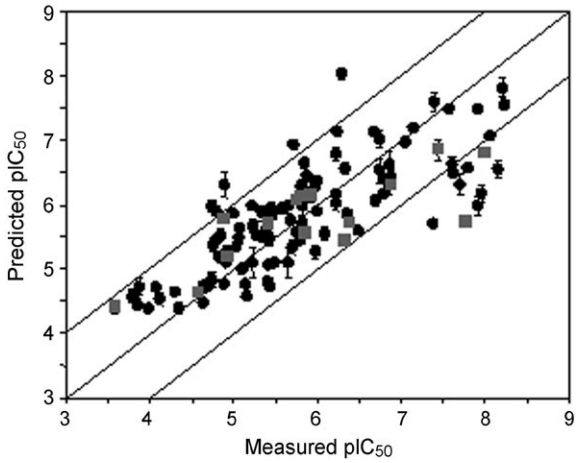

Figure 4. Mean measured and predicted pIC_{50} values for the training, test, and validation set for the conservatively selected descriptor set including standard deviation (PLS model: shape, EALmin, Naryl, shape^2, and shape^3 descriptors, split according to pIC_{50} bins); ● = test and training set, ■ = validation set. Mean R^2y(train) = 0.64, N = 75; mean R^2y(test) = 0.61, N = 23; mean RMSE(train&test) = 0.68; mean MUE(train&test) = 0.52, N = 98; mean R^2y(validation) = 0.62; mean RMSE(validation) = 0.79; mean MUE(validation) = 0.63, N = 15.

Naryl indicates the number of aromatic rings present. It enters all 20 regression equations with a positive sign, suggesting that hERG blockade potency increases with the number of aromatic rings. The shape descriptor indicates the similarity of the molecular shape to astemizole, cisapride, and sertindole. It is calculated as an autocorrelation graph of surface points and compared with the corresponding graph of the target molecules. All three shape descriptors enter the equations with positive signs, suggesting that the more similar the shape of the compounds from this dataset to those of the most active compounds, the lower their observed IC_{50} values. EALmin is an unexpected descriptor, as all the EAL descriptors are related to interactions in which the molecule acts as an electron acceptor. If donor–acceptor interactions in which the substrate is the acceptor play a significant role, one would expect EALmax to appear. More positive EALmax values indicate a strong capacity to undergo electron-accepting interactions, as in hydrogen bonding, for example. EALmin enters the regression equation with a positive sign, suggesting that the more negative the minimum electron affinity, the fewer molecules tend to block hERG. Areas of the most negative local electron affinity are often located over carbonyl groups, if present (see Figure 2c, blue area). Over the whole dataset, EALmin covers a range of approximately −150 to −100 kcal mol^{-1}. Areas of low local electron affinity associated with carbonyls lie between approximately −150 to −120, compared with −125 to −100 for compounds without carbonyl groups. It is known that carbonyls reduce the hERG affinity.[55] Thus, we tried replacing the EALmin descriptor with a 1/0 descriptor for the presence/absence of carbonyl groups. However, this descriptor was never selected in the MLR approach for any of the 20 training sets. Thus, EALmin apparently represents a descriptor that enables us to go one step deeper into the analysis of carbonyls for hERG blockade by describing the exact electronic nature of the carbonyl more precisely.

Pharmacophore-driven models

When we compared models 1 and 2 (and other models that are not shown), we observed that four potent hERG blockers were always predicted to have far lower pIC_{50} values: sertindole analogue SA15,[9] clemastine, tolterodine, and haloperidol. All four have high measured and lower predicted pIC_{50} values, and are shown in Figure 5.

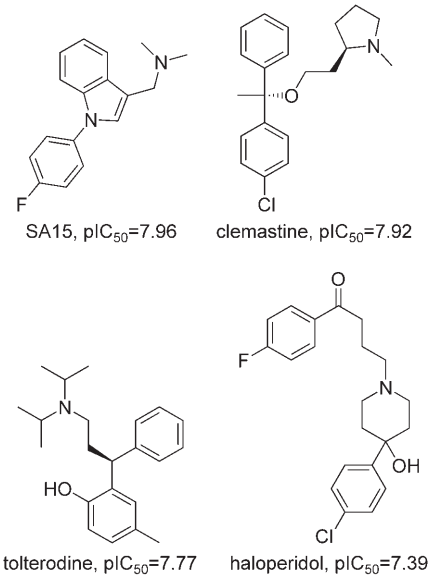

Figure 5. Consistent outliers in the general models.

The first three molecules have fairly clear similar chemical structures, whereas haloperidol is not obviously similar. As they all have very high pIC_{50} values, most of their molecular features probably participate in binding. A pharmacophore model (in the following denoted as "pharmacophore 1") shown in Figure 6 was constructed from these compounds. Haloperidol fits into this pharmacophore in a U-shaped conformation to give a fairly high internal Catalyst fit value of 2.96 (with each of the features ranked one in case of a perfect overlap).

As only 113 compounds were available for regression models, we decided to analyze the pharmacophores using the whole set to include as much information as possible. Because of the small number of data openly available, patterns found here must be validated properly with independent validation sets as soon as more data become available. In the following, we use the phrase test set in place of validation set, as the set of compounds not used in training is no longer strictly completely independent of the model-building procedure; it has already seen the pharmacophore built from outliers of the pri-

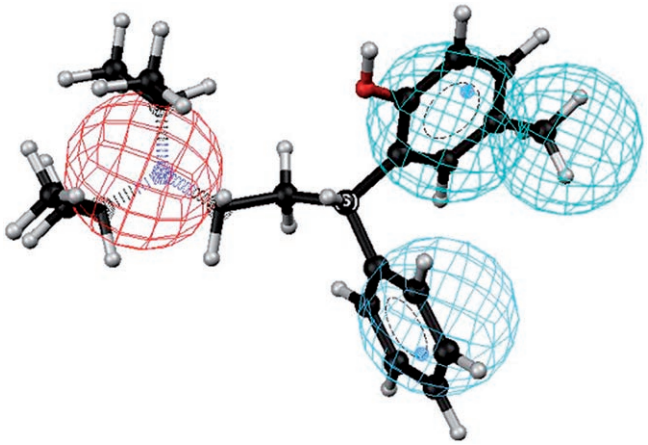

Figure 6. Pharmacophore 1 with docked tolterodine; red = positively ionizable nitrogen, blue = aromatic hydrophobic feature, turquoise = hydrophobic feature.

mary models. For the following models, we included Clog*P* as a further descriptor, because it is known that lipophilicity plays a role in hERG binding.[47]

Model 3.1: unspecific binding, pharmacophore non-hitters

Pharmacophore 1 represents a subgroup of previously published pharmacophores, a minimum pharmacophore for hERG blockade. We overlaid the whole set of 113 molecules with pharmacophore 1; 51 compounds hit the pharmacophore, 62 compounds did not. Molecules that hit the pharmacophore have, on average, higher pIC_{50} values than those that do not. For creating a new model, 30% of the compounds that did not hit the pharmacophore were partitioned off randomly to form a test set. A PLS model was built for the corresponding training and test sets using the MLR_Opt feature-selection algorithm on the training set.

Only EALmin, Naryl, shape, and shapeQ2 entered the QSAR equation of model 3.1. This is a noteworthy result, as we expected many more descriptors to enter this model. Moreover, compounds from this group seem to be dominant in determining model 2 described above for the entire dataset.

For the molecules that hit pharmacophore 1, a test set was split off, and a PLS model was generated. However, this model performed poorly relative to that for the molecules that did not hit the pharmacophore (details not shown). There is no significant dependence between hERG blockade and descriptors in this descriptor and compound set.

To include more specific knowledge about the hERG receptor, we built a second pharmacophore from astemizole, cisapride, flunarizine, and sertindole. These compounds show the lowest IC_{50} values published. Thus, a pharmacophore derived from them can be expected to illustrate the spatial distribution of important features. All of the compounds hit pharmacophore 1. The sertindole analogues from the study of Pearlstein et al.,[9] which also have high pIC_{50} values, were excluded in order to avoid overemphasis of the influence of the sertindole scaffold. This procedure yields a pharmacophore that is very similar to pharmacophore 1. Instead of a hydrophobic feature in close proximity to an aromatic hydrophobic hot spot, it contains a hydrophobic feature that is almost collinear with the positive nitrogen atom and the other hydrophobic feature. This model is, for instance, very similar to that published by Cavalli et al.[49] Pharmacophore 2 and compounds from which it was built are shown in Figure 7.

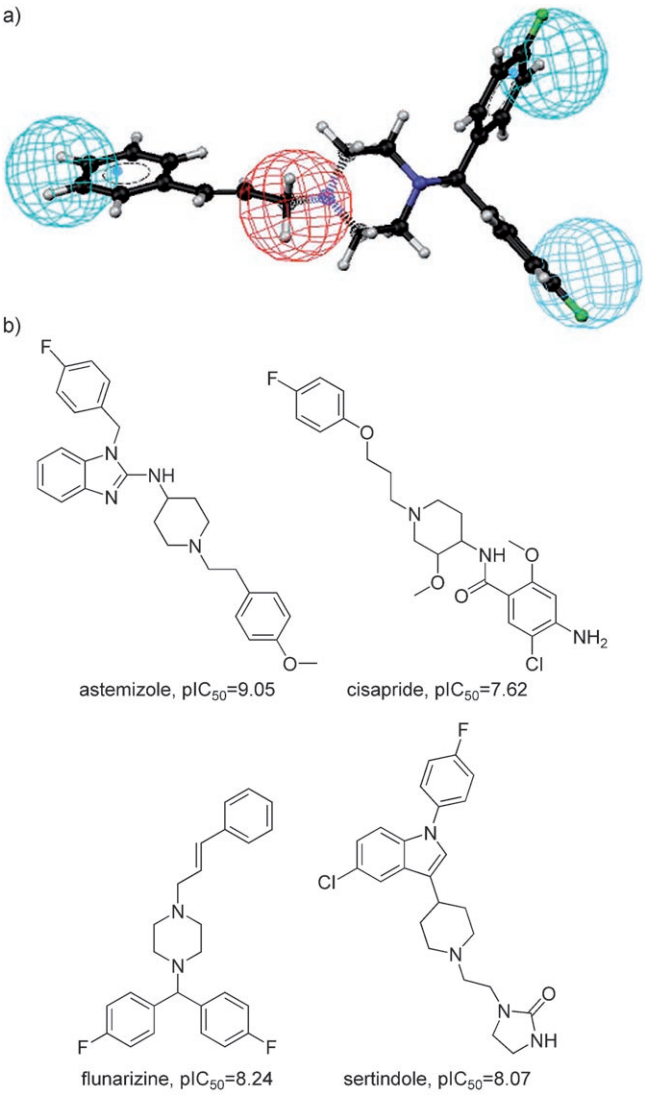

Figure 7. a) Pharmacophore 2 with docked flunarizine; red = positively ionizable nitrogen, blue = aromatic hydrophobic feature, turquoise = hydrophobic feature; b) strongest hERG blockers reported.

Model 3.2: specific binding type 2, pharmacophore 2

In order to differentiate between molecules that hit the minimum pharmacophore and those that hit this more specific pharmacophore, the whole dataset was searched with pharmacophore 2; 20 compounds hit each feature of the pharmacophore, and 17 of these compounds also hit each feature of pharmacophore 1. Six of these 20 compounds were selected randomly as test set. A PLS model was created based on the remaining 14 compounds as training set. Descriptors selected

for this model with the MLR_Opt approach are globularity, POLmin, SHANIbar, SHANIvar and Ndon. Molecules that display all the features of pharmacophore 2 have an average pIC_{50} of 7.11.

Model 3.3: specific binding type 1, pharmacophore 1

From the residual set of 34 molecules that hit pharmacophore 1 but not pharmacophore 2, 11 were split off randomly as a test set. A PLS model from the 23 training set molecules was generated using the MLR_Opt feature selection; esHBac, VppQ2, and Clog*P* were selected as relevant descriptors.

A plot of measured vs. predicted pIC_{50} values for the whole dataset after partitioning according to pharmacophores (i.e. using models 3.1–3.3 according to whether the compound hits either none or at least one of the pharmacophores 1 or 2) is shown in Figure 8. Statistical measures for models 1 and 3 are

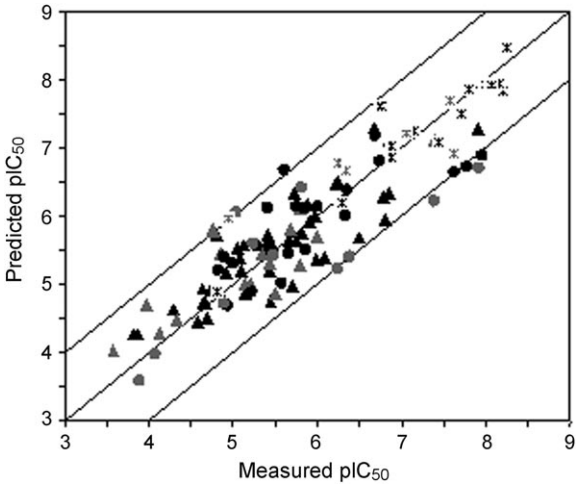

Figure 8. Overall measured and predicted pIC_{50} values for the training and test set for the MLR_Opt descriptor set, calculated according to pharmacophore hits (PLS model: complete prediction, random splits); ▼ = test set non-hitters, ▲ = training set non-hitters, ● = training set pharmacophore 1 hitters, ● = test set pharmacophore 1 hitters, * = training set pharmacophore 2 hitters, * = test set pharmacophore 2 hitters. R^2y(train) = 0.82, RMSE(train) = 0.45, MUE(train) = 0.35, N = 79; R^2y(test) = 0.72, RMSE(test) = 0.61, MUE(test) = 0.49, N = 34.

summarized in Table 1. A flow chart of the procedure of splitting according to pharmacophore hits is shown in Figure 9.

Model 4: most significant descriptors of composite models

To analyze the significance of different descriptors for each individual sub-model, the inclusion criterion for descriptors, the *F* value, in the stepwise MLR procedure was changed as for

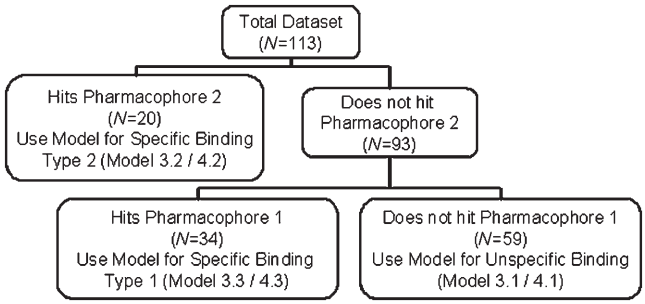

Figure 9. Flowchart of the composite model.

model 2. For each model, $F = 10$ turned out to be a good exclusion criterion for any random descriptor. All three subsets were randomly split into 70% training and 30% test data 20 times. The stepwise MLR procedure was carried out up to five times with exclusion of previously selected descriptors for each of the 60 training sets.

Model 4.1

For the 20 training sets of the group of compounds that do not hit any pharmacophore, EALmin, Naryl, shape, and shapeQ2 were identified as significant because they were selected between 16 and 20 times. Shape is more than 90% correlated (R^2) with shapeQ2. As we later generated PLS models designed to circumvent the multi-collinearity problem, both were used. EALbar and EALbar− were also selected 18 and 19 times, respectively. However, in 90% of the cases they were selected after EALmin, which is correlated between 60–80% with EALbar and EALbar−. Thus, we reasoned that EALmin is a better descriptor for this dataset than EALbar or EALbar− and excluded the last two. Whereas shape and shapeQ2 were selected with equal frequency, shapeQ1, which is also highly correlated ($R^2 > 0.9$) with shape and shapeQ2, is selected after shape and shapeQ2 in all but one case. Thus, shapeQ1 was excluded. All other descriptors were selected far less frequently. The final set of relevant descriptors here is EALmin, Naryl, shape, and shapeQ2.

Table 1. Statistical properties for the composite PLS models.

		Model 3.1 No Pharmacophore	Model 3.2 Pharmacophore 2	Model 3.3 Pharmacophore 1	Model 1 General Model
Descriptors		EALmin, Naryl, shape, shapeQ2	POLmin, SHANIbar, SHANIvar, Ndon	esHBac, VppQ2m, Clog*P*	EALmax, EALmin, POLmin, SHANIbar, Naryl, shapeQ2, shapeQ4
R^2y	training	0.74	0.90	0.63	0.81
	validation	0.69	0.79	0.64	0.68/0.70
RMSE	training	0.43	0.34	0.57	0.50
	validation	0.51	0.57	0.77	0.74
MUE	training	0.34	0.21	0.44	0.37
	validation	0.41	0.48	0.63	0.55
N		59	20	34	113

Model 4.2

For the 20 training sets of pharmacophore 2 hitters, QsumH was selected 19 times, and Ndon and POLmin were selected 17 times. All the other descriptors were selected far less frequently.

Model 4.3

For the 20 training sets of compounds that hit exclusively pharmacophore 1, only Clog *P* turned out to be significant, as it was selected 19 of 20 times in the first step of MLR analysis. All other descriptors were selected only up to eight times.

For all the 60 sets, PLS models based on the relevant descriptors were built. The mean model performances including the validation set predictions are summarized in Table 2.

Table 2. Mean statistical properties for PLS composite and general models with conservative descriptor selection.

		Model 4.1 No Pharmacophore	Model 4.2 Pharmacophore 2	Model 4.3 Pharmacophore 1	Model 4.1–4.3 All[a]	Model 2 General Model[b]
Descriptors		EALmin, Naryl, shape, shapeQ2	Ndon, QsumH, POLmin	Clog *P*	Dependent on type of binding	EALmin, Naryl, shape, shape^2, shape^3
R^2y	training	0.74	0.84	0.41	0.67	0.63
	test	0.66	0.76	0.39		0.62
RMSE	training	0.46	0.40	0.80	0.56	0.68
	test	0.49	0.46	0.83		0.79
MUE	training	0.35	0.29	0.61	0.43	0.52
	test	0.40	0.38	0.67		0.63
N	training	42	14	24	80	98
	test	17	6	10	33	15

[a] Test set values only (from 20-fold cross-validation). [b] Training set + test set = training, validation set = test.

We have investigated descriptor-based QSAR equations with and without partitioning according to two different pharmacophores. We find that partitioning improves the predictivity of the overall hERG model, although this in itself is not surprising, as we have generated three local QSAR equations out of one general equation. The literature indicates that hERG channel blockers can be accommodated in different binding modes within the pore.[9,23,30,49,56] It is tempting to interpret our composite model as reflecting three different binding modes. This is a dangerous conclusion because it rests entirely on the improved fitting obtained with the composite model. We have, for instance, identified three compounds (ajmaline, amsacrine, and domperidone) that fall in our nonspecific model, but are known[57–59] to cause QT prolongation. Any conclusions based on the assignment of a given model are therefore untenable at the moment. This is, however, partly because the biological data on the compounds that actually cause QT prolongation are not available. We can hope that once such data become available, our model can be connected to physiological effects, but this is currently not possible. Therefore, we consider the improved performance of a composite consisting of three different models as no more than a hint towards three different types of binding, two of which exhibit individual sets of important pharmacophoric features, and the third, less specific binding that depends only on shape, and not specific pharmacophoric features. However, our composite model fits well into the spirit of ensemble-based systems in decision making[60] and can claim some physical justification because of the known tendency of hERG toward multiple binding modes. The use of pharmacophores as selection tools for descriptor-based models avoids the need for scoring functions, with all their associated problems.

Specific binding type 2 (Pharmacophore 2, Model 3.2/4.2)

The best-known hERG binding pattern consists of one positively ionizable feature and three aromatic hydrophobic features in the periphery around the positive center. Cavalli et al.,[49] Pearlstein et al.,[9] and Ekins et al.[50] have all described this feature previously. In our model 3.2, all molecules that exhibit all these features could be predicted very accurately with the first component of a PLS model, giving an R^2y value around 0.80. Once all pharmacophore features are present, hERG blockade depends on the minimum of the local polarizability, which is usually associated with the most positively charged hydrogen atom, the sum of the electrostatic potential on all hydrogen atoms, and the number of H-bond donors, as demonstrated in model 4, which only uses the most significant descriptors. EALmin and the number of H-bond donors are clearly related to the strength and number of H-bonds, respectively, in which the substrate is the donor. In all 20 PLS models there is only one vector that contains all three descriptors with almost equal loadings after mean centering and adjusting the standard deviation to one.

The number of H-bond donors correlates negatively with pIC_{50}, which may indicate that the desolvation penalty for H-bond donors cannot be compensated by any of the H-bond-accepting features of the hERG channel associated with this type of specific binding. For this set of compounds we found neither Clog *P* nor any of the initial descriptors to be significantly correlated with the pIC_{50} of hERG blockade.

Specific binding type 1 (Pharmacophore 1, Model 3.3/4.3)

Roughly 50% of the total dataset hit pharmacophore 1. This pharmacophore consists of one positively ionizable feature, one hydrophobic feature, and one aromatic hydrophobic feature with a hydrophobic feature in close proximity. The arrangement is very similar to parts of pharmacophore 2, but pharmacophore 2 has one less hydrophobic feature. Seventeen

of 20 molecules that hit pharmacophore 2 also hit pharmacophore 1. A separate PLS model was built for all the molecules that hit pharmacophore 1 exclusively. In the single test–training set split, we found Clog *P*, esHBac, and VppQ2 for this set. In the conservative descriptor selection of model 4, we only found Clog *P* to be significant for this set. It enters the model with a positive coefficient, indicating that increased lipophilicity will lead to an increased hERG blockade for compounds from this group.

In a docking study, Farid et al.[23] found a completely different orientation for SA15 inside the pore relative to sertindole. This finding is consistent with the two binding types suggested by this study.

Nonspecific binding (Non-hitters, Model 3.1/4.1)

A third PLS model was built for the remaining molecules that do not hit either of the two pharmacophores. Descriptors for this model are EALmin, Naryl, shape, and shapeQ2, selected by both the MLR_Opt selection algorithm and the conservative descriptor selection of model 6. Using only these four descriptors, around 70% of the variation in pIC_{50} values can be explained. The PLS equation for model 3.1 after mean centering and adjusting the standard deviation to one is given below.

$$pIC_{50\,unspecific} = 6.51 + 0.52\,EALmin + 0.32\,Naryl + 0.33\,shape + 0.29\,shapeQ2 \quad (1)$$

We interpret this model as being appropriate for molecules that bind nonspecifically to the hERG channel. The model probably covers a wide range of different binding modes and perhaps several binding sites as well. This type of binding can perhaps be compared with binding to serum protein and should be amenable to a nonspecific, QSPR-like model.[61] However, some indications can be found that the model does relate specifically to hERG. High shape similarity to the most active compounds astemizole, cisapride, and sertindole leads to an increase in the calculated hERG-blocking effect. Furthermore, the number of aromatic rings and the minimum of the electron affinity have an effect on hERG blockade properties. These descriptors are consistent with the idea of an unspecific binding mode of some hERG blockers that depends more on hydrophobic and van der Waals interactions than any specific binding features. This nonspecific binding model applies to compounds that have a lower effect on hERG blockade than specific binders (those that hit one of the two pharmacophores). All compounds from the pharmacophore groups are predicted to have lower or similar pIC_{50} values by the unspecific model than by the appropriate specific model. Nonspecific binding probably occurs throughout the part of the pore that is directed toward the interior of the cell. Experimentally, at least 28 (7×4) amino acids in the pore can contribute to ligand binding (see "Important structural features" above). The descriptors for the QSAR model of unspecific binders reflect this situation.

Alternative pharmacophores

The pharmacophores identified herein are very similar to those proposed in previous studies. It has been suggested earlier[29,30,33] that the aromatic and hydrophobic interactions probably take place with Phe656 and Tyr652. The role of the positively ionizable nitrogen is not yet clear. It has been suggested that it undergoes cation–π interactions with Tyr652.[9,44,50,56] Recently, Farid et al. found that by docking potent hERG blockers into the open form of the channel, the ionizable nitrogen atoms are too far away from Tyr652 to make reasonable contributions to the binding energy.[23] Despite the lack of a clear role, the ionizable nitrogen is an important feature of almost all high-affinity pharmacophores that have been described in the literature. However, Aronov recently described a pharmacophore for hERG blockers without the positively ionizable nitrogen feature.[52] We were unable to find compounds with very high reported hERG pIC_{50} values that hit this pharmacophore. The highest IC_{50} value that was published for a neutral compound is 1.9 μM for ketoconazole. Thus, it is not necessary to consider this pharmacophore in our work because the corresponding compounds would be assigned to the group of unspecific binders.

Conclusions

We have presented a composite model for hERG blockade that differentiates between nonspecific binding and two types of specific binding. The prediction quality of this model exceeds that of our own general models, leading to a flow chart of how the pharmacophore and QSAR models can be applied in practice.

We used predominantly the ParaSurf[62] set of descriptors, which consists of quantum mechanically derived descriptors based on surface properties, with additional basic 1D and 2D descriptors. These descriptors have proven to be appropriate for the final QSAR models and provide some insight into the key molecular features that drive hERG blocking potential in each of the subgroups.

The technique of partitioning compounds according to a pharmacophore comparison and then partitioning the dataset for a series of individual descriptor-based QSAR equations represents a useful alternative to pharmacophore similarities, scoring functions, etc., for all problems in which multiple binding modes may play a role, that is, for promiscuous targets such as hERG or cytochromes P450. More data for hERG blockers may reveal further pharmacophores, which will then give some more specific QSAR equations.

Experimental Section

Data collection

All publicly available hERG blockade IC_{50} data were collected in June 2006. Fenichel's list,[63] previously published papers, and PubMed[64] were used as starting points for data collection. IC_{50} values were collected only if at least an abstract of the original work was available. Personally communicated values were exclud-

ed. IC_{50} values were collected from the original publications to ensure the highest possible quality of the dataset and to avoid transcription errors. This procedure gave 345 experimental values for 206 compounds.

Data preparation

For model generation, only measurements on hERG channels cloned into human embryonic kidney (HEK-293) and Chinese hamster ovary (CHO) cells were used to ensure consistency in the dataset. Measurements on other mammalian cell lines such as *Cercopithecus aethiops* (COS-7) and atrial tumor (AT-1) were excluded. Measurements on *Xenopus* oocytes (XO) were excluded. It is well known that measurements on XO yield IC_{50} values that are on average 12-fold greater than those measured on mammalian cell lines, because XO membranes are more lipophilic than those of CHO and HEK-293 cells.[56] Measurements on Purkinje fibers were excluded because these native cells differ even more from HEK-293 and CHO cells than other mammalian cell lines.

A radioligand replacement method was introduced for medium- to high-throughput determination of hERG inhibition. [^{3}H]Dofetilide[65] or [^{3}H]astemizole[66] are replaced competitively from mammalian cells that overexpress hERG channels. However, this is not a functional test, and not every compound that replaces dofetilide must necessarily exhibit the same blocking properties. Furthermore, as the pore is quite large, some compounds could bind allosterically to dofetilide. We have therefore not considered these measurements in assembling our dataset.

Our literature study revealed that some published fits of the Hill equation[67] for high-range IC_{50} values are not statistically significant[68] (see oleandomycin). Hence, we reviewed all reported IC_{50} values greater than 20 μM. We considered at least one measured concentration above 75% inhibition to be necessary for the IC_{50} value to be included. This 75% criterion should ensure reliable fitting.

Table 3. Descriptors used in this study.

	Molecular Electrostatic Potential Descriptors[73]
V_{max} (MEPmax)	Maximum (most positive) MEP value
V_{min} (MEPmin)	Minimum (most negative) MEP value
V_+(mean) (meanMEP+)	Mean of the positive MEP values
V_-(mean) (meanMEP−)	Mean of the negative MEP values
V(mean) (meanMEP)	Mean of all MEP values
ΔV (MEP-range)	MEP range
σ^2_+ (MEPvar+)	Total variance in the positive MEP values
σ^2_- (MEPvar−)	Total variance in the negative MEP values
σ^2_{tot} (MEPvartot)	Total variance in the MEP
ν (MEPbalance)	MEP balance parameter
$\sigma^2_{tot}\nu$ (var_balance)	Product of the total variance in the MEP and the MEP parameter
	Local Ionization Energy Descriptors[73]
IE_L^{max} (IELmax)	Maximum value of the local ionization energy
IE_L^{min} (IELmin)	Minimum value of the local ionization energy
IE_L(mean) (IELbar)	Mean value of the local ionization energy
ΔIE_L (IEL-range)	Range of the local ionization energy
σ^2_{IE} (IELvar)	Variance of the local ionization energy
	Local Electron Affinity Descriptors[73]
EA_L^{max} (EALmax)	Maximum of the local electron affinity
EA_L^{min} (EALmin)	Minimum of the local electron affinity
EA_{L+}(mean) (EALbar+)	Mean of the positive values of the local electron affinity
EA_{L-}(mean) (EALbar−)	Mean of the negative values of the local electron affinity
EA_L(mean) (EALbar)	Mean value of the local electron affinity
ΔEA_L (EAL-range)	Range of the local electron affinity
σ^2_{EA+} (EALvar+)	Variance in the local electron affinity for all positive values
σ^2_{EA-} (EALvar−)	Variance in the local electron affinity for all negative values
σ^2_{EAtot} (EALvartot)	Sum of the positive and negative variances in the local electron affinity
ν_{EA} (EALbalance)	Local electron affinity balance parameter
δA^+_{EA} (EALfraction+)	Fraction of the surface area with positive local electron affinity
	Mean Local Electronegativity[73]
χ_L(mean) (ENEGbar)	Mean value of the local electronegativity
	Local Polarizability Descriptors[73]
α_L^{max} (POLmax)	Maximum value of the local polarizability
α_L^{min} (POLmin)	Minimum value of the local polarizability
α_L(mean) (POLbar)	Mean value of the local polarizability
$\Delta\alpha_L$ (POL-range)	Range of the local polarizability
σ^2_α (POLvar)	Variance in the local polarizability
	Shannon Entropy Descriptors[72]
H_{int}^{max} (SHANImax)	Maximum value of the internal Shannon entropy
H_{ext}^{max} (SHANEmax)	Maximum value of the external Shannon entropy
H_{int}^{min} (SHANImin)	Minimum value of the internal Shannon entropy
H_{ext}^{min} (SHANEmax)	Minimum value of the external Shannon entropy
H_{int}(mean) (SHANIbar)	Mean value of the internal Shannon entropy
H_{ext}(mean) (SHANEbar)	Mean value of the external Shannon entropy
σ_{int}^2 (SHANIvar)	Variance in the internal Shannon entropy
σ_{ext}^2 (SHANEvar)	Variance in the external Shannon entropy
H_{int}^{mol} (SHANItot)	Internal molecular Shannon entropy
H_{ext}^{mol} (SHANEtot)	External molecular Shannon entropy
	Autocorrelation Descriptors[a]
shape	Shape autocorrelation
shapeQ(1-4)	Shape autocorrelation for quartal 1–4 (2.5–4.36 Å; 4.42–6.28 Å; 6.34–8.2 Å; 8.26–10.12 Å)
Vpp	Plus–plus MEP autocorrelation
VppQ(1–4)	Plus–plus MEP autocorrelation for quartal 1–4
Vmm	Minus-minus MEP autocorrelation
VmmQ(1–4)	Minus–minus MEP autocorrelation for quartal 1–4
Vpm	Plus–minus MEP autocorrelation
VpmQ(1–4)	Plus–minus MEP autocorrelation for quartal 1–4

Table 3. (Continued)

Additional Descriptors	
μ (dipole)	Dipole moment
μ_D (dipden)	Dipolar density
A (polarizability)	Molecular electronic polarizability
M_w	Molecular weight
G (globularity)	Globularity
A (totalarea)	Molecular surface area
VOL (volume)	Molecular volume
Qsum	Sum of the VESPA[75] electrostatic potential on all (N, O, P, S, F, Cl, Br, I, H, hal) atoms
Estate	Analogous to the Kier & Hall Estate[76] using the bond order between atom i and j instead of the distance
Estate2	Analogous to the Kier & Hall Estate using r_{ij} to describe the distance between atom i and j
LocPol	Local polarity: all absolute deviations from the mean ESP for each surface point summed up and divided by the number of surface points
CovHBac	Covalent hydrogen bond acidity: E_{HOMO}(molecule)−E_{LUMO}(water)
CovHBbas	Covalent hydrogen bond basicity: E_{LUMO}(molecule)−E_{HOMO}(water)
esHBac	Electrostatic hydrogen bond acidity: most negative formal charge (molecule)−most positive formal charge (water)
esHBbas	Electrostatic hydrogen bond basicity: most positive formal charge on hydrogen (molecule)−most negative formal charge (water)
CohIndex	Cohesive index: (Nacc × Ndon$^{0.5}$)/total surface
Nacc	Number of H-bond acceptors
Ndon	Number of H-bond donors
Naryl	Number of aromatic rings
Clog P	Calculated octanol/water distribution coefficient (using MOE[45])

[a] Autocorrelation similarities were calculated as described in the ParaSurf'06 manual.[62] Astemizole, cisapride, and sertindole were used as reference points because they show the highest hERG blockade.

For our regression models, only compounds with an exact reported IC_{50} value were used. All calculations used pIC_{50} values.

4-Aminopyridine and chlorobutanol were excluded from the dataset because of their extremely small size. Metoprolol was excluded because it has been shown that, in contrast to all other compounds tested, it does not lose binding affinity when Phe 656 or Tyr 652 are mutated.[35] It must therefore bind in a different mode than the other compounds. For compounds with multiple IC_{50} values, the mean IC_{50} value was used if $pIC_{50\,max} - pIC_{50\,min} < 1$; otherwise the compounds were excluded.

Overall, we obtained 113 compounds that spanned a range from about $pIC_{50} = 3.7$ to $pIC_{50} = 9$. The complete dataset is listed in the Supporting Information.

The dataset was split into training, test, and validation sets. First, the compounds were partitioned into six pIC_{50} bins of equal range, from which a validation set of 15 compounds was selected randomly. The pIC_{50} distribution of the validation set was thus similar to that of the whole set. From the remaining 98 compounds, test sets of 23 compounds with distributions similar to that of the 98 compounds were picked randomly out of the pIC_{50} bins. The random selection was repeated 20 times.

Descriptor calculation

All calculations were performed with the uncharged forms of the molecules, because charges in an unknown binding mode inside an unknown binding pocket cannot be inferred from charges appropriate to physiological pH and because protonatable nitrogen atoms are directly exchangeable for protonated ones within QSAR and pharmacophore models. After generating 3D coordinates with CORINA,[69] the structures were optimized in the gas phase with AM1.[70,71] From the resulting geometries, statistical variables that describe the distribution of the local molecular electrostatic potential (MEP), the local ionization energy (IEL), the local electron affinity (EAL), the local polarizability (POL), as well as external and internal Shannon entropies[72] (SHANI, SHANE) on the molecular surface were calculated.[73] Furthermore, shape and charge surface autocorrelations and some basic 1D and 2D descriptors were calculated. Altogether this gave a set of 99 descriptors, which is listed in Table 3. All descriptors were calculated using the program ParaSurf'06.[62] For later refinements after pharmacophore subdivision, Clog P[74] was added.

Statistical modeling

MLR calculations were performed with TSAR 3.3.[41] For PLS analysis we used SIMCA P-11.[77] LibSVM[78] was used for the ε-SVR computations. ε-SVR parameters were determined during 50 steps of a simplex optimization of the cross-validation performance.[79] This procedure is more efficient than a systematic grid search. Pharmacophore modeling and searching was done with Catalyst.[80]

Descriptor selection

Descriptors of model 1 were selected based on the 98-membered training and test set after removal of the validation set. Descriptors of models 3.1–3.3 were selected based on the corresponding training sets. Two different algorithms were used. In the first approach, descriptors were selected by stepwise MLR based on the whole descriptor set. In a second approach, the stepwise MLR algorithm was refined. Stepwise MLR only extracts one set of descriptors and always follows the steepest gradient. It does not necessarily extract the best set of descriptors. Thus, we ran three MLRs, each time removing the previously selected descriptors from the initial set. From the combined set of all descriptors selected by MLR and their squares, we excluded those that lowered or did not change R^2_{cv} in PLS models. This procedure yielded the optimized MLR set of descriptors (MLR_Opt). Unsupervised forward selection (UFS),[81] a formal inference-based recursive modeling (FIRM)[82–84] decision tree implemented in TSAR 3.3, and variable influence on projection (VIP) statistics in PLS models were also investigated as alternative approaches to select descriptors. However, the resulting models performed similarly or worse than the models described, and thus are not discussed further herein.

Pharmacophore generation

Pharmacophores 1 and 2 were generated using default Catalyst parameters (Catalyst 4.10).[80]

Acknowledgements

We thank Kerstin Höhfeld (University of Erlangen–Nürnberg) for preparing the surface pictures. C.K. thanks Boehringer–Ingelheim Pharma GmbH & Co. KG for financial support. We thank Cepos InSilico Ltd. for providing us with the ParaSurf'06 program.

Keywords: hERG · ion channels · semi-empirical calculations · structure–activity relationships · toxicology

[1] J. W. Warmke, B. Ganetzky, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 3438–3442.
[2] F. De Ponti, E. Poluzzi, A. Cavalli, M. Recanatini, N. Montanaro, *Drug Saf.* **2002**, *25*, 263–286.
[3] T. Meyer, K. H. Boven, E. Gunther, M. Fejtl, *Drug Saf.* **2004**, *27*, 763–772.
[4] W. J. Crumb, I. Cavero, *Pharm. Sci. Technol. Today* **1999**, *2*, 270–280.
[5] M. L. Roy, R. Dumaine, A. M. Brown, *Circulation* **1996**, *94*, 817–823.
[6] H. Suessbrich, S. Waldegger, F. Lang, A. E. Busch, *FEBS Lett.* **1996**, *385*, 77–80.
[7] Z. Zhou, V. R. Vorperian, Q. Gong, S. Zhang, C. T. January, *J. Cardiovasc. Electrophysiol.* **1999**, *10*, 836–843.
[8] M. L. Roy, A. Dennis, A. M. Brown, *FEBS Lett.* **1997**, *417*, 28–32.
[9] R. A. Pearlstein, R. J. Vaz, J. Kang, M. Preobrazhenskaya, A. E. Shchekotikhin, A. M. Korolev, L. N. Lysenkova, O. V. Miroshnikova, J. Hendrix, D. Rampe, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1829–1835.
[10] F. De Ponti, E. Poluzzi, N. Montanaro, *Eur. J. Clin. Pharmacol.* **2001**, *57*, 185–209.
[11] M. L. De Bruin, M. Petterson, R. H. Meyboom, A. W. Hoes, H. G. Leufkens, *Eur. Heart J.* **2005**, *26*, 590–597.
[12] M. C. Sanguinetti, J. S. Mitcheson, *Trends Pharmacol. Sci.* **2005**, *26*, 119–124.
[13] S. Viskin, *Lancet* **1999**, *354*, 1625–1633.
[14] M. E. Curran, I. Splawski, K. W. Timothy, G. M. Vincent, E. D. Green, M. T. Keating, *Cell* **1995**, *80*, 795–804.
[15] M. C. Sanguinetti, C. Jiang, M. E. Curran, M. T. Keating, *Cell* **1995**, *81*, 299–307.
[16] M. Trudeau, J. W. Warmke, B. Ganetzky, G. A. Robertson, *Science* **1995**, *269*, 92–95.
[17] M. C. Sanguinetti, M. Tristani-Firouzi, *Nature* **2006**, *440*, 463–469.
[18] P. B. Bennett, K. Yazawa, N. Makita, A. L. George, *Nature* **1995**, *376*, 683–685.
[19] M. Recanatini, E. Poluzzi, M. Masetti, A. Cavalli, F. De Ponti, *Med. Res. Rev.* **2005**, *25*, 133–166.
[20] A. M. Brown, D. Rampe, *Pharm. News* **2000**, *7*, 15–20.
[21] Y. A. Kuryshev, E. Ficker, L. Wang, P. Hawryluk, A. T. Dennis, B. A. Wible, A. M. Brown, J. Kang, X.-L. Chen, K. Sawamura, W. Reynolds, D. Rampe, *J. Pharmacol. Exp. Ther.* **2005**, *312*, 316–323.
[22] R. Rajamani, B. A. Tounge, J. Li, C. H. Reynolds, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1737–1741.
[23] R. Farid, T. Day, R. A. Friesner, R. A. Pearlstein, *Bioorg. Med. Chem.* **2006**, *14*, 3160–3173.
[24] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, R. MacKinnon, *Science* **1998**, *280*, 69–77.
[25] Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, R. Mackinnon, *Nature* **2002**, *417*, 515–522.
[26] S. B. Long, E. B. Campbell, R. MacKinnon, *Science* **2005**, *309*, 897–903.
[27] Y. Zhou, J. H. Morais-Cabral, A. Kaufman, R. MacKinnon, *Nature* **2001**, *414*, 43–48.
[28] Y. Jiang, A. Lee, J. Chen, M. Cadene, B. T. Chait, R. MacKinnon, *Nature* **2002**, *417*, 523–526.
[29] J. S. Mitcheson, J. Chen, M. Lin, C. Culberson, M. C. Sanguinetti, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 12329–12333.
[30] D. Fernandez, A. Ghanta, G. W. Kauffman, M. C. Sanguinetti, *J. Biol. Chem.* **2004**, *279*, 10120–10127.
[31] J. P. Lees-Miller, Y. Duan, G. Q. Teng, H. J. Duff, *Mol. Pharmacol.* **2000**, *57*, 367–374.
[32] J. A. Sanchez-Chapula, T. Ferrer, R. A. Navarro-Polanco, M. C. Sanguinetti, *Mol. Pharmacol.* **2003**, *63*, 1051–1058.
[33] J. Sanchez-Chapula, R. A. Navarro-Polanco, M. C. Sanguinetti, *J. Biol. Chem.* **2002**, *277*, 23587–23595.
[34] K. Kamiya, R. Niwa, J. S. Mitcheson, M. C. Sanguinetti, *Mol. Pharmacol.* **2006**, *69*, 1709–1716.
[35] K. Kawakami, T. Nagatomo, H. Abe, K. Kikuchi, H. Takemasa, B. D. Anson, B. P. Delisle, C. T. January, Y. Nakashima, *Br. J. Pharmacol.* **2006**, *147*, 642–652.
[36] R. Pearlstein, R. Vaz, D. Rampe, *J. Med. Chem.* **2003**, *46*, 2017–2022.
[37] F. Osterberg, J. Aqvist, *FEBS Lett.* **2005**, *579*, 2939–2944.
[38] L. Du, M. Li, Q. You, L. Xia, *Biochem. Biophys. Res. Commun.* **2007**, *355*, 889–894.
[39] O. Roche, G. Trube, J. Zuegge, P. Pflimlin, A. Alanine, G. Schneider, *ChemBioChem* **2002**, *3*, 455–459.
[40] *Molecular Operating Environment (MOE)* version 2004.03, Chemical Computing Group Inc., Montreal, QC, Canada, **2004**.
[41] *TSAR* version 3.3, Accelrys Software Inc., Cambridge, UK, **2003**.
[42] M. Seierstad, D. K. Agrafiotis, *Chem. Biol. Drug Des.* **2006**, *67*, 284–296.
[43] C. L. Gavaghan, C. H. Arnby, N. Blomberg, G. Strandlund, S. Boyer, *J. Comput.-Aided Mol. Des.* **2007**, *21*, 189–206.
[44] G. Cianchetta, Y. Li, J. Kang, D. Rampe, A. Fravolini, G. Cruciani, R. J. Vaz, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3637–3642.
[45] A. M. Aronov, *Drug Discovery Today* **2005**, *10*, 149–155.
[46] A. M. Aronov, B. B. Goldman, *Bioorg. Med. Chem.* **2004**, *12*, 2307–2315.
[47] M. J. Waring, C. A. Johnstone, *Bioorg. Med. Chem. Lett.* **2007**, *17*, 1759–1764.
[48] M. Song, M. Clark, *J. Chem. Inf. Comput. Sci.* **2006**, *46*, 392–400.
[49] A. Cavalli, E. Poluzzi, F. De Ponti, M. Recanatini, *J. Med. Chem.* **2002**, *45*, 3844–3853.
[50] S. Ekins, W. J. Crumb, R. D. Sarazan, J. H. Wikel, S. A. Wrighton, *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427–434.
[51] L. P. Du, K. C. Tsai, M. Y. Li, Q. D. You, L. Xia, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4771–4777.
[52] A. M. Aronov, *J. Med. Chem.* **2006**, *49*, 6917–6921.
[53] M. K. Leong, *Chem. Res. Toxicol.* **2007**, *20*, 217–226.
[54] D. J. Livingstone, D. W. Salt, *J. Med. Chem.* **2005**, *48*, 661–663.
[55] S. R. Johnson, H. Yue, M. L. Conder, H. Shi, A. M. Doweyko, J. Lloyd, P. Levesque, *Bioorg. Med. Chem.* **2007**, *15*, 6182–6192.
[56] W. Bains, A. Basman, C. White, *Prog. Biophys. Mol. Biol.* **2004**, *86*, 205–233.
[57] W. Haverkamp, G. Moennig, P. Kirchhof, L. Eckardt, M. Borggrefe, G. Breithardt, *Z. Kardiol.* **2001**, *90*, 586–590.
[58] C. L. Schwartz, K. S. Bender, P. J. Burke, J. S. Kan, C. I. Civin, *Cancer Treat. Rep.* **1984**, *68*, 1043–1044.
[59] B. Drolet, G. Rousseau, P. Daleau, R. Cardinal, J. Turgeon, *Circulation* **2000**, *102*, 1883–1885.
[60] R. Polikar, *IEEE Circuits and Systems Magazine* **2006**, *3*(06), 21–45.
[61] J. R. Votano, M. Parham, L. M. Hall, L. H. Hall, L. B. Kier, S. Oloff, A. Tropsha, *J. Med. Chem.* **2006**, *49*, 7169–7181.
[62] T. Clark, A. H. C. Horn, J.-H. Lin, *ParaSurf'06*, Cepos InSilico Ltd., Ryde, UK, **2006**; http://www.ceposinsilico.com.
[63] R. Fenichel, http://www.fenichel.net/pages/Professional/subpages/QT/Tables/pbydrug.htm.
[64] National Institutes of Health (NIH), http://www.ncbi.nlm.nih.gov.
[65] K. Finlayson, L. Turnbull, C. T. January, J. Sharkey, J. S. Kelly, *Eur. J. Pharmacol.* **2001**, *430*, 147–148.
[66] P. J. Chiu, K. F. Marcoe, S. F. Bounds, C. H. Lin, J. J. Feng, A. Lin, F. C. Cheng, W. J. Crumb, R. Mitchell, *J. Pharmacol. Sci.* **2004**, *95*, 311–319.
[67] A. V. Hill, *J. Physiol.* **1910**, *40*, 9–12.
[68] W. Volberg, B. J. Koci, W. Su, J. Lin, J. Zhou, *J. Pharmacol. Exp. Ther.* **2002**, *302*, 320–327.
[69] J. Gasteiger, *CORINA* version 3.3; Molecular Networks Inc., Erlangen, Germany, **2006**.
[70] T. Clark, A. Alex, B. Beck, F. Burkhardt, J. Chandrasekhar, P. Gedeck, A. H. C. Horn, M. Hutter, B. Martin, G. Rauhut, W. Sauer, T. Schindler, T. Steinke, *VAMP* version 8.2, Accelrys Software Inc., San Diego, CA, USA, **2002**.
[71] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P. Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[72] K. G. Byler, T. Clark, M. J. de Groot in *Proceedings of the Beilstein Workshop "Molecular Interactions: Bringing Chemistry to Life"*, Frankfurt, Germany, **2006**.

[73] B. Ehresmann, M. J. de Groot, A. Alex, T. Clark, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 658–668.
[74] *Dayprop*, Daylight Chemical Information Inc., Aliso Viejo, CA, USA, **2006**.
[75] B. Beck, R. C. Glen, T. Clark, *J. Comput. Chem.* **1997**, *18*, 744–756.
[76] L. B. Kier, L. H. Hall, *Molecular Structure Description*, Academic Press, San Diego, **1999**.
[77] *SIMCA P-11*, Umetrics AB, Umeå, Sweden, **2005**.
[78] C. C. Chang, C. J. Lin, LIBSVM: a library for support vector machines, http://www.csie.ntu.edu.tw/~cjlin/libsvm.
[79] U. Norinder, *Neurocomputing* **2003**, *55*, 337–346.
[80] *Catalyst Software Package* version 4.10, Accelrys Software Inc., San Diego, CA, USA, **2003**.
[81] D. C. Whitley, M. G. Ford, D. J. Livingstone, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1160–1168.
[82] D. M. Hawkins, S. S. Young, A. R. Ill, *Quant. Struct.-Act. Relat.* **1997**, *16*, 296–302.
[83] S. S. Young, D. M. Hawkins, *J. Med. Chem.* **1995**, *38*, 2784–2788.
[84] S. S. Young, D. M. Hawkins, *SAR QSAR Environ. Res.* **1998**, *8*, 183–193.

Received: August 24, 2007
Revised: October 22, 2007
Published online on December 5, 2007

