

# A Virtual Screening Method for Prediction of the hERG Potassium Channel Liability of Compound Libraries

Olivier Roche, Gerhard Trube, Jochen Zuegge, Pascal Pflimlin, Alexander Alanine, and Gisbert Schneider<sup>\*[a,b]</sup>

*A computer-based method has been developed for prediction of the hERG (human ether-à-go-go related gene) K<sup>+</sup>-channel affinity of low molecular weight compounds. hERG channel blockage is a major concern in drug design, as such blocking agents can cause sudden cardiac death. Various techniques were applied to finding appropriate molecular descriptors for modeling structure–activity relationships: substructure analysis, self-organizing maps (SOM), principal component analysis (PCA), partial least squares fitting (PLS), and supervised neural networks. The most accurate pre-*

*diction system was based on an artificial neural network. In a validation study, 93% of the nonblocking agents and 71% of the hERG channel blockers were correctly classified. This virtual screening method can be used for general compound-library shaping and combinatorial library design.*

## KEYWORDS:

bioinformatics · combinatorial chemistry · drug design · neural network · structure–activity relationships

## Introduction

The K<sup>+</sup> channel encoded by the human ether-à-go-go related gene (hERG or KCNH2) gives rise to the rapid component of the delayed rectifier K<sup>+</sup>-channel current, I<sub>Kr</sub>. The hERG K<sup>+</sup> channel plays a crucial role for normal action potential repolarization in the heart. It has been used as a therapeutic target for class-III anti-arrhythmic agents, but a wide range of noncardiac drugs also inhibit the hERG K<sup>+</sup> channel, resulting in a drug-induced long QT syndrome (LQTS) that can cause sudden cardiac death.<sup>[1]</sup> This potentially lethal side effect is a major issue for the development of any new drug, since it has been shown that various molecules, such as antihistamines, psychoactive agents, calcium antagonists, and antimicrobials, can inhibit the hERG K<sup>+</sup> channel.<sup>[2]</sup> It is therefore important to assess the hERG blocking potential of novel chemical structures as early as possible during the drug discovery process. The only precise method for determination of the compounds' potencies in K<sup>+</sup>-channel inhibition, is patch-clamp electrophysiology, which is time-consuming and labor-intensive. Higher-throughput methods, such as radioligand binding or detection of membrane potential changes by fluorescent dyes, are indirect and not very reliable. With the aid of a large set of compounds tested by patch-clamping over several years of research, we have now developed a fast, computer-based, virtual screening method for the prediction of the hERG blocking potential of new compounds.

## Results and Discussion

### Data generation and compilation

The experimental data generation process has been performed over the past three years. A standardized patch-clamp procedure

was used to compile a unique homogenous data set containing hERG K<sup>+</sup>-channel inhibition data for 472 compounds. The aim was to establish a structure–activity relationship (SAR) model. We grouped the data set into three classes:

Class 1: 96 compounds with IC<sub>50</sub> < 1 μM (low IC<sub>50</sub>),

Class 2: 148 compounds with IC<sub>50</sub> > 10 μM (high IC<sub>50</sub>), and

Class 3: 228 compounds with 1 μM ≤ IC<sub>50</sub> ≤ 10 μM (medium IC<sub>50</sub>).

Class 1 included some potent hERG K<sup>+</sup>-channel blockers known from the literature and retested in our laboratory: cisapride, dofetilide, E-4031, haloperidole, and terfenadine.

The first approach was to consider all available information by using all three classes of compounds for SAR modeling. Unfortunately, this approach completely failed to produce a useful scoring scheme. This is most probably due to the error range of the experimental results (up to twofold), which tends to give rise to an "ill-posed" problem; molecules with similar features are classified into different classes (low, medium, high). To avoid this problem, we followed the "likeness concept", which only uses the extremes of the data set (high and low IC<sub>50</sub> classes); this resulted in 244 nonredundant compounds.<sup>[3]</sup>

[a] Dr. G. Schneider, Dr. O. Roche, G. Trube, Dipl.-Biol. J. Zuegge, P. Pflimlin, Dr. A. Alanine  
F. Hoffmann-La Roche AG  
Pharmaceuticals Division  
4070 Basel (Switzerland)  
Fax: (+41) 6168-87408  
E-mail: gisbert.schneider@roche.com

[b] Present address:  
Dr. G. Schneider  
Johann Wolfgang Goethe University  
Institute of Organic Chemistry  
Marie-Curie-Strasse 11, 60439 Frankfurt/Main (Germany)  
E-mail: gisbert.schneider@modlab.de

For model validation we compiled from the literature a set of 38 drugs with known hERG K<sup>+</sup>-channel IC<sub>50</sub> values. As the experimental conditions for measuring IC<sub>50</sub> values for these drugs can be quite diverse, possibly resulting in significant

### Descriptor generation and selection

A large number of different descriptor types was generated in order to capture relevant molecule features. We first analyzed general properties of the data (Table 2). The only significant difference was found for the calculated octanol/water partition coefficient computed by the clogP routine.<sup>[5]</sup> Blocking molecules (low IC<sub>50</sub>) tend to be more lipophilic than nonblocking agents. It should be stressed that increasing lipophilicity usually increases binding to protein receptors, which is a general phenomenon not restricted to the hERG channel. This feature alone is not sufficient for classification, so additional sets of descriptors were computed:

- One-dimensional (1D): 120 atom types defined by Ghose and Crippen (GC descriptors),<sup>[6b]</sup> 78 TSAR descriptors,<sup>[7]</sup> and pKa<sup>[8]</sup>
- Two-dimensional (2D): 150 topological CATS descriptors<sup>[9]</sup>
- Three-dimensional (3D): 56 VolSurf descriptors<sup>[10]</sup>
- 853 DRAGON descriptors including the BCUT, WHIM, 2D-autocorrelation, 3D-Morse, and Getaway descriptors<sup>[11]</sup>

**Table 1.** IC<sub>50</sub> values and prediction scores of 38 drugs.

Name	IC <sub>50</sub> [μM]	Score	Name	IC <sub>50</sub> [μM]	Score
Astemizole	0.001	0.03	Loratadine	0.17	0.80
Azimilide	0.6	0.78	Mdl74156	12.1	1.00
Bepidil	0.55	0.01	Mexiletine	≥ 10	0.92
Ciprofloxacin	966	0.94	Mizolastine	0.35	0.76
Clarithromycin	720	1.00	Mk-499	0.032	1.00
Demethylastemizol	0.001	0.01	Moxifloxacin	129	0.85
Diltiazem	17	1.00	Nifedipine	≥ 50	1.00
Domperidone	0.16	0.05	Nitrendipine	≥ 10	1.00
Droperidol	0.032	0.03	Olanzapine	0.2	0.42
Ebastine	0.14	0.01	Ondansetron	0.81	0.63
Erythromycin	≥ 10	1.00	Pimozide	0.018	0.01
Gatifloxacin	130	0.88	Risperidone	0.14	0.02
Glibenclamide	74	0.93	Sertindole	0.014	0.74
Glimepiride	> 500	0.82	Sildenafil	30	0.15
Grepafloxacin	44	0.79	Sparfloxacin	18	0.78
Halofantrine	0.2	0.05	Sulfamethoxazole	10000	0.91
Isradipine	≥ 10	0.96	Trimethoprim	240	0.98
Ketoconazole	49	1.00	Verapamil	0.8	0.02
Levofloxacin	915	0.81	Ziprasidone	0.15	0.22

differences in the results, these values were only used as an independent validating set, and not for elaboration of the prediction scheme (Table 1). An additional set of 57 compounds, which originated from other patch-clamp experiments performed at Roche, was also used for validation.

Several techniques were applied to the identification of appropriate molecular descriptors for SAR modeling: substructure analysis, self-organizing maps (SOM), principal component analysis (PCA), partial least squares fitting (PLS), and supervised neural networks.

### Substructure analysis

With the aid of the commercially available software tool LeadScope (release 2RC1),<sup>[4]</sup> we tried to identify substructures capable of discriminating between hERG K<sup>+</sup>-channel blockers (Class 1, low IC<sub>50</sub>) and nonblockers (Class 2, high IC<sub>50</sub>). LeadScope describes compounds in terms of approximately 27000 predefined structural features and displays their distribution by means of separate histograms. Prevalent structural elements that might be characteristic of the data sets were manually selected from the program output. Strikingly, we were not able to identify individual substructures exclusively present in either one of the extremes. Nevertheless, some weak trends were revealed: pairs of hydrogen-bond donors separated by six bonds were present in 57% of the nonblocking agents and 30% of the blockers, and benzenesulfonyl groups were found in 20% of the nonblockers but in only 2% of the blockers. It must be stressed that 1-R-4-alkyl-benzene moieties (where R = any atom) were also found in 49% of the blockers. This observation suggests a general structural bias in the data rather than individual substructure differences that are meaningful characteristic features of hERG K<sup>+</sup>-channel blockers or nonblockers. A more extensive analysis based on additional molecular descriptors was therefore required.

**Table 2.** Properties of the training data.

Properties	hERG K <sup>+</sup> -channel affinity	
	high IC <sub>50</sub> <sup>[a]</sup>	low IC <sub>50</sub> <sup>[a]</sup>
molecular weight	333 (72)	345 (45)
lipophilicity (clogP) <sup>[5]</sup>	2.70 (1.16)	4.28 (1.12)
number of hydrogen-bond donors	1.68 (1.11)	1.47 (0.75)
number of hydrogen-bond acceptors	4.96 (1.79)	3.59 (1.24)
drug-likeness score <sup>[3]</sup>	0.71 (0.24)	0.77 (0.20)
"frequent-hitter" score <sup>[3]</sup>	0.18 (0.29)	0.11 (0.19)

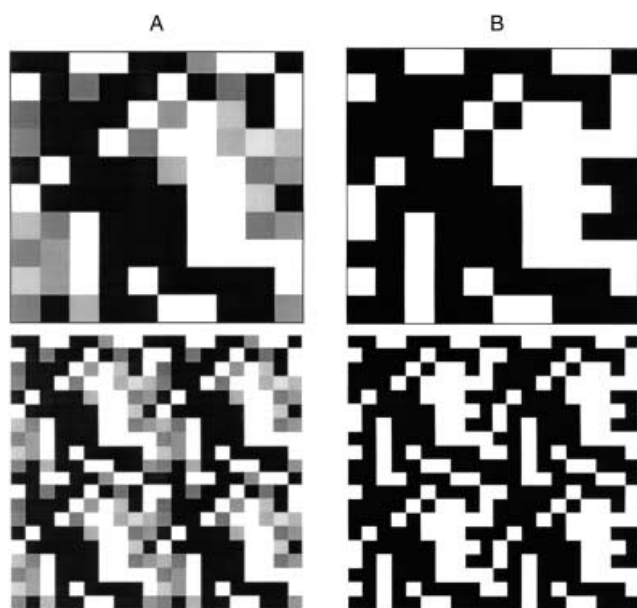
[a] Values in parentheses are the standard deviations.

For the three-dimensional descriptors, the molecular conformations were generated and optimized with the programs Corina and Cosmic respectively, as part of the TSAR 3.21 program.<sup>[7]</sup> In total, 1258 descriptors were generated. The SOM technique was applied to identify a relevant descriptor subset. The SOM approach generates a topology-preserving nonlinear mapping of a high-dimensional space to a low-dimensional space.<sup>[12]</sup> In this case the 1258-dimensional descriptor space was projected onto the plane, and the standard Matthews correlation coefficient for binary data, *cc*, was used to estimate the classification ability of the map [Eq. (1)].<sup>[12, 13]</sup>

$$cc = \frac{(NP) - (OU)}{\sqrt{(N+O)(N+U)(P+O)(P+U)}} \quad (1)$$

In Equation (1), *N*, *P*, *O*, and *U* are the numbers of true negative (low IC<sub>50</sub>; Class 1 compounds), true positive (high IC<sub>50</sub>; Class 2 compounds), false positive, and false negative predictions, respectively. A perfect prediction gives a correlation coefficient of 1.

The highest classification accuracy (*cc* = 0.70) was observed for the VolSurf and GC descriptors (Figure 1). The other descriptors gave values of: CATS: *cc* = 0.68, WHIM: *cc* = 0.61, Getaway: *cc* = 0.57, 3D-Morse: *cc* = 0.60, all 853 DRAGON descriptors: *cc* = 0.58. For reasons of simplicity and calculation speed, the GC descriptors were kept for further study.



**Figure 1.** Self-organizing map (SOM) projection of the compound distribution in a high-dimensional space spanned by 120 GC descriptors.  $10 \times 10$  clusters were formed. A) Density of high  $IC_{50}$  compounds (white = none, black = many). B) Binary classification of chemical space; high  $IC_{50}$  compounds area in black, low  $IC_{50}$  compounds area in white. The map forms a torus, as illustrated by the  $(2 \times 2)$  replicas of the maps (lower row). The map does not contain empty neurons.

### Linear prediction model

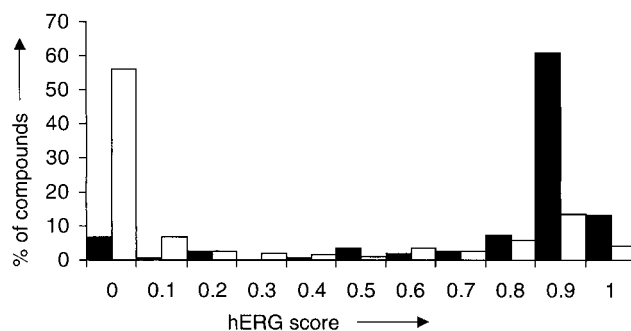
Having found a useful molecular representation for the given prediction task, PCA was performed to extract a small set of orthogonal factors describing the data distribution as defined by the raw GC descriptors.<sup>[14]</sup> From the PCA, seven outliers were identified (six nonblockers and one blocker). After this step, a PLS model—a multivariate linear regression technique—was elaborated to obtain a first prediction scheme. The SIMCA-P software package was used for this purpose.<sup>[15]</sup> To estimate the stability of these models to outliers, we compared the ranking of the descriptors extracted with the variable influence on projection parameters (VIP) values computed by SIMCA-P with respect to the removal of the outliers identified by PCA. Variable ranking was not significantly affected by outliers, a fact which suggests that the solution may not be critically influenced by extreme raw data values. The best linear prediction tool derived from the PLS correctly reclassified 84% of the nonblocking compounds, but only 77% of the blocking agents, thereby yielding a binary Matthews correlation coefficient of  $cc_{\text{training}} = 0.61$ . This value is low for reclassification, which indicates that a linear model might not be appropriate for discriminating between potent hERG  $K^+$  blockers and nonblockers on the basis of the particular descriptor set chosen.

### Nonlinear prediction model

The next step was a nonlinear prediction method, as given by a supervised neural network system. Three-layered supervised neural networks were used to find a discriminating scheme between hERG  $K^+$ -channel blockers and nonblockers. Such

systems are universal function estimators suited for quantitative structure–activity relationship (QSAR) modeling.<sup>[16]</sup> In our case, their architecture contained an input layer (fan-out units), one hidden layer (sigmoidal units), and a single sigmoidal output unit.<sup>[17]</sup> These networks were trained by an evolutionary algorithm implementing adaptive step-size control, as detailed elsewhere.<sup>[16]</sup> The mean-square error served as the objective function that had to be minimized during the network training. In different training runs, the number of hidden layer units and generations was systematically varied to find an appropriate setting. The desired output value (target value) of the neural network was 0 for compounds with a low  $IC_{50}$  value (Class 1) and 1 for compounds with a high  $IC_{50}$  value (Class 2). Ten-times cross-validation was performed with random 80% (training) and 20% (testing) splits of the data. For further validation, a validation set, containing 95 compounds, was classified. The “overlearning” (or “overtraining”) effect is often a main concern during neural network training. To avoid this effect, the process was terminated when the  $cc$  value reached an optimum for both the test set and the validation set (forced stop). In addition, the least complex network was selected as the final prediction model.

Of the networks tested, that with two hidden neurons seemed to be best suited. This network reached a Matthews correlation of  $cc_{\text{training}} = 0.85$  (reclassification of the training set; 93% correct) and an average Matthews correlation of  $cc_{\text{test}} = 0.61$  for the ten random testing sets, which means that 89% of the nonblocking compounds and 70% of the blocking agents were correctly classified. Figure 2 shows the distribution of the raw prediction scores for the test sets produced by this network. These results indicate that the prediction tool is able to distinguish between compounds with extreme properties (low and high  $IC_{50}$ ) but is more suited to identify nonblocking compounds.



**Figure 2.** Distribution of the neural network output values (“hERG score”) for the test data (ten-times cross-validation result). The target values were 1 for high  $IC_{50}$  compounds (black bars), and 0 for low  $IC_{50}$  compounds (white bars). The y axis gives the fractions of test compounds receiving a certain score value.

The validation set contained 72 compounds with high  $IC_{50}$  values and only 23 compounds with low  $IC_{50}$  values. 93% of the nonblockers and 71% of the blockers were correctly predicted, to yield a Matthews correlation of  $cc_{\text{validation}} = 0.66$ . This result demonstrates that the prediction model is flexible enough to make acceptable predictions for a diverse set of molecules.

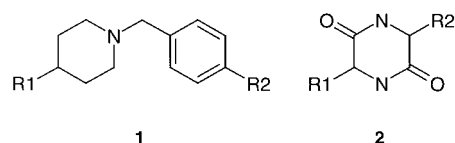
To assess the limits of the prediction scheme, we analyzed the false positives and false negatives produced for both the training

and the validating sets. Five drugs from the validation set were misclassified as nonblockers (Scheme 1). Other known blockers such as dofetilide, E-4031, cisapride, astemizole, and verapamil were correctly categorized. For virtual screening purposes, it is particularly desirable not to miss hERG channel blocking agents. Therefore, the misclassification of known blockers reveal some particular deficits of our prediction model. Very probably the molecular descriptors used here do not appropriately represent all essential features required for perfect prediction. This also points to a general disadvantage in the neural network method, namely the fact that an easy understanding of the decisive molecular features can hardly be obtained. Nevertheless, the prediction scheme is able to identify features shared by most of the compounds of each class, as indicated by its good overall performance. It identifies a trend towards hERG K<sup>+</sup>-channel blocking activity rather than accurate predictions for each molecule. As a consequence, it shows a moderate capacity to distinguish small structural differences.

An explanation for misclassification can also be found in the structure of the data set. Since the data set is project-related, the

diversity of the compounds is limited. Moreover, structural motifs shared by many known hERG K<sup>+</sup>-channel blocking compounds, such as “aromatic–linker–aromatic” (where the linker consists of 2–10 bonds, usually including a basic nitrogen), are underrepresented in the data set. This is because medicinal chemists usually try to avoid synthesis of such molecules. Moreover, the different inhibitors might bind to different receptor sites and affect the channel through different mechanisms. Generally speaking, the data set contains less information about the blockers than about the nonblockers. This is reflected by the better performance on molecules that do not inhibit the hERG K<sup>+</sup> channel.

Two virtual combinatorial libraries were analyzed by our prediction method to demonstrate its applicability to compound library shaping. Scaffold 1 represents a structural motif found in

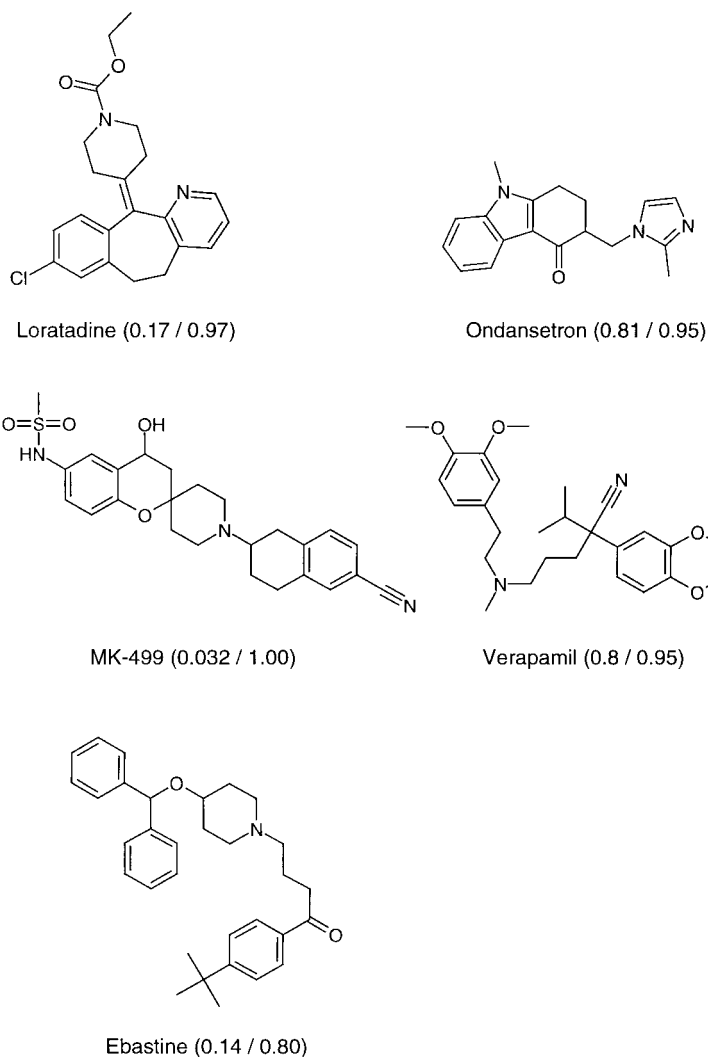


many known hERG K<sup>+</sup>-channel blockers. Scaffold 2 was designed to be devoid of any such known motif. Virtual libraries were enumerated by attaching the identical set of 100 generic representative building blocks to each scaffold exit vector (R<sup>1</sup>, R<sup>2</sup>), resulting in 10 000 virtual products. From the prediction results, the library based on scaffold 1 contains 58% potential hERG K<sup>+</sup>-channel inhibitors, whereas the library of compounds based on 2 contains only 0.1% potential blockers. This result was expected and clearly shows that compound libraries, scaffolds, and building block selections can be ranked on the basis of the prediction score. It must be stressed that although a prediction may not be perfect for an individual compound, general trends are accurately recognized and therefore the prediction model qualifies for virtual library shaping.

The new prediction scheme will join the already available suite of *in silico* filters for drug-likeness, frequent hitters, cytotoxicity, bioavailability, and others.<sup>[18]</sup> It can be used as a general filter for hERG K<sup>+</sup>-channel liability to prioritize compound collections from large databases and to design new virtual libraries. The work presented in this paper should be considered as a baseline study for assessing hERG affinity *in silico*. Since new experimental results are constantly being produced and added to the model, we are convinced that the prediction tool can be improved further.

## Experimental Section

The cell line stably expressing the hERG K<sup>+</sup> channel was obtained from GENION (Hamburg). The complementary DNA coding for the human ether-à-go-go related gene product (hERG; GenBank access no. U04270) had been cloned into the pcDNA3 vector (Invitrogen, Carlsbad, California). Plasmids had been introduced into CHO cells by using DMRIE-C (Gibco, Carlsbad, California) as a transfection agent. Cells were grown in minimal essential

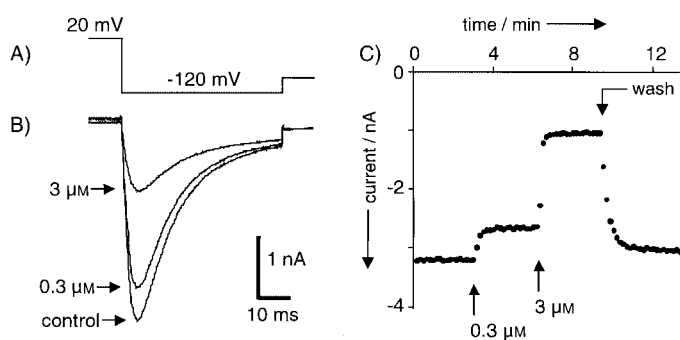


**Scheme 1.** Structures of five misclassified drugs from the validation set. Notation: name (IC<sub>50</sub> (μM)/score).

medium alpha (MEM; Gibco, Carlsbad, California) supplemented with 10% (v/v) heat-inactivated fetal calf serum, 1% (v/v) penicillin/streptomycin/glutamine solution (Gibco, Carlsbad, California), and 1 mg/ml G418 (geneticine). For electrophysiology experiments the cells were continuously superfused by extracellular saline containing 150 mM NaCl, 10 mM KCl, 1 mM MgCl<sub>2</sub>, 3 mM CaCl<sub>2</sub>, and 10 mM HEPES. Compounds for testing were dissolved in dimethylsulfoxide at a concentration of 10 mM and diluted in extracellular saline to the desired concentration (mostly 1 and 10 μM). The glass micropipettes for whole-cell patch-clamp recording were filled with intracellular saline containing 110 mM KCl, 4.5 mM MgCl<sub>2</sub>, 10 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES), 10 mM 1,2-bis(2-Aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid, 4 mM Na<sub>2</sub>ATP, 20 mM Na<sub>2</sub>-creatine-phosphate, and 0.2 g L<sup>-1</sup> creatine-kinase.

The whole-cell configuration of the patch-clamp technique was used for current recording.<sup>[19]</sup> Cells were clamped to a holding potential of -80 mV, and the hERG K<sup>+</sup> channels were activated once every 10 s by a voltage pulse pattern consisting of a 1 s depolarization to 20 mV followed by a 20 ms hyperpolarization to -120 mV.<sup>[20]</sup> The amplitude of the transient inward current at -120 mV was used for further analysis (Figure 3A, B). Current responses were recorded for about three minutes under control conditions. The test compound was then applied to the investigated cell from a nearby capillary, usually for another three minutes per concentration. Current amplitude values were plotted versus time to depict the effect of the compound (Figure 3C) and calculate the percentage of channel inhibition (y). We tried to test at least two concentrations (c), usually differing by a factor of ten, and bracketing the 50% inhibitory concentration (IC<sub>50</sub>). The results obtained from at least three cells were pooled and fitted by Equation (2) to estimate the mean values of IC<sub>50</sub>, the 95% confidence limits, and the Hill coefficient *H*.

$$y = \frac{100\%}{1 + 10^{H(\log IC_{50} - \log c)}} \quad (2)$$



**Figure 3.** Example of an experiment showing the effect of quinidine on the membrane current in a Chinese hamster ovary (CHO) cell expressing the hERG K<sup>+</sup> channel. A) Part of the voltage pulse pattern. B) Superimposed current signals evoked by the voltage pulses before drug application (control) and in the presence of 0.3 and 3 μM quinidine. C) Plot of the peak current amplitudes versus time. The onset of drug application (0.3 and 3 μM) and its end (wash) are indicated by the arrows.

Since the logarithms of *c* and IC<sub>50</sub> are used in the fitting equation, the upper and lower confidence limits of IC<sub>50</sub> are not equidistant from the mean but must be expressed as a multiple or fraction of the mean value. In 95% of cases, the confidence interval was narrower than the range from half to twice the mean. The average value of *H* was 0.90 (standard deviation 0.17). If the lowest (*a*) or highest (*b*) tested concentration caused more or less than 50% inhibition, respectively, we did not attempt a fit, but concluded that the IC<sub>50</sub> was below *a* or larger than *b*.

Hans-Joachim Böhm, Eric Ertel, Holger Fischer, Petra Schneider, Manfred Kansy, and Page Mahaney are thanked for many valuable discussions and support.

- [1] a) M. T. Keating, M. C. Sanghinetti, *Cell* **2001**, *104*, 569–580; b) M. C. Trudeau, J. W. Warmke, B. Ganetzky, G. A. Robertson, *Science* **1995**, *269*, 92–95.
- [2] J. I. Vandenberg, B. D. Walker, T. J. Campbell, *Trends Pharmacol. Sci.* **2001**, *22*, 241–246.
- [3] a) J. Sadowski, H. Kubinyi, *J. Med. Chem.* **1998**, *41*, 3325–3329; b) O. Roche, P. Schneider, J. Zuegge, W. Guba, M. Kansy, A. Alanine, K. Bleicher, F. Danel, E.-V. Gutknecht, M. Rogers-Evans, W. Neidhart, H. Stalder, M. Dillon, E. Sjogran, N. Fotouhi, P. Gillespie, R. Goodnow, W. Harris, P. Jones, M. Taniguchi, S. Tsujii, W. van der Saal, G. Zimmermann, G. Schneider, *J. Med. Chem.* **2002**, *45*, 137–142.
- [4] G. Roberts, G. J. Myatt, W. P. Johnson, K. P. Cross, P. E. Blower, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1302–1214.
- [5] W. M. Meylan, P. H. Howard, *J. Pharm. Sci.* **1995**, *84*, 83–92.
- [6] a) V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, R. K. Robins, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 163–172; b) A. K. Ghose, G. M. Crippen, *J. Comput. Chem.* **1986**, *7*, 565–577.
- [7] TSAR 3.21, Oxford Molecular Ltd, The Medawar Center, Oxford Science Park, Oxford, OX4 4GA (UK).
- [8] pKalc 3.21, CompuDrug Chemistry Ltd., PO Box 405, 1395 Budapest 62 (Hungary).
- [9] G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem.* **1999**, *111*, 3068–3070; *Angew. Chem. Int. Ed.* **1999**, *38*, 2894–2895.
- [10] VolSurf v 2.0.6, Multivariate Infometric Analysis S.r.l., Viale dei Catagni 16, 06143 Perugia (Italy).
- [11] a) R. Todeschini, V. Consonni, *Handbook of molecular descriptors*, Wiley-VCH, Weinheim, **2000**; b) DRAGON v1.11, Milano Chemometrics and QSAR group, P.za della Scienza 1, 20126 Milano (Italy).
- [12] a) T. Kohonen, *Biol. Cybern.* **1982**, *43*, 59–69; b) M.-L. Lee, G. Schneider, *J. Comb. Chem.* **2001**, *3*, 284–289.
- [13] B. W. Matthews, *Biochim. Biophys. Acta* **1975**, *405*, 442–451.
- [14] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS)*, Umeå:Umetrix, **1999**.
- [15] a) S. Wold, *Chemometrics Intell Lab Systems*, **1994**, *23*, 149–161; b) SIMCA-P 8.0, Umetrics AB, Box 7960, 907 19 Umeå (Sweden).
- [16] G. Schneider, P. Wrede, *Prog. Biophys. Mol. Biol.* **1998**, *70*, 175–222.
- [17] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to The Theory of Neural Computation*, Addison-Wesley, Redwood City, **1991**.
- [18] H.-J. Boehm, G. Schneider (Eds.), *Virtual Screening for Bioactive Molecules*, Wiley-VCH, Weinheim, **2000**.
- [19] O. P. Hamill, A. Marty, E. Neher, B. Sakmann, F. J. Sigworth, *Pflügers Arch.* **1981**, *391*, 85–100.
- [20] P. L. Smith, T. Baukowitz, G. Yellen, *Nature* **1996**, *379*, 833–836.

Received: December 10, 2001 [F 329]