Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 3637–3642
Bioorganic & Medicinal Chemistry Letters

Predictive models for hERG potassium channel blockers

Giovanni Cianchetta,[a,b] Yi Li,[a] Jiesheng Kang,[a] David Rampe,[a] Arnaldo Fravolini,[b] Gabriele Cruciani[c] and Roy J. Vaz[a,*]

[a]*Sanofi-Aventis Pharmaceuticals, 1041 Route 202/206 N, Bridgewater, NJ 08807, USA*
[b]*Dipartimento di Chimica e Tecnologia del Farmaco, Università di Perugia, Via del Liceo 1, I-06123 Perugia, Italy*
[c]*Dipartimento di Chimica, Laboratorio di Chemiometria, Università di Perugia, Via Elce di Sotto, 10, I-06123 Perugia, Italy, and Molecular Discovery Ltd., 4, Chandos Street, W1A 3BQ, London, United Kingdom*

Received 17 December 2004; accepted 14 March 2005

Abstract—We report here a general method for the prediction of hERG potassium channel blockers using computational models generated from correlation analyses of a large dataset and pharmacophore-based GRIND descriptors. These 3D-QSAR models are compared favorably with other traditional and chemometric based HQSAR methods.

The risk factor of QT prolongation associated with drug inhibitory interaction with human ether-a-go-go (hERG) potassium channels has been linked to drug induced sudden cardiac death and represents an important safety concern by pharmaceutical developers and regulators.[1–4] A sound strategy has been adopted to screen out hERG channel blockers at the early stage of drug discovery using available in vitro methods[5] such as the Rb-efflux functional test,[6,7] radio-labeled binding assays,[8,9] or the whole cell patch-clamp assay,[10–12] which provides the most reliable data measuring drug interactions with the hERG potassium channels. In addition, numerous efforts have been pursued to devise a fast in silico screen using computational models to predict binding affinities of hERG channel blockers.[13]

Several computational models have been reported recently for hERG channel blockers. Most of these models were derived from ligand-based approaches using a relatively small experimental data set containing molecules of positively charged functional group(s). Using a training set of 15 molecules, a pharmacophore model generated by CATALYST consisted of four hydrophobic features and one positive ionizable feature.[14] Similarly, a pharmacophore model established from CoMFA analysis of a set of 31 QT-prolonging drugs involved a positively charged tertiary amine flanked by three aromatic or hydrophobic centers.[15] We proposed a 'drain plug' model from a combined homology modeling of the hERG channel pore region and a ligand-based 3D-QSAR CoMSiA analysis of an experimentally consistent dataset.[16]

Meanwhile, chemometric approaches have been developed for binary classifications in order to discriminate hERG channel blockers from a library of compounds. Training of artificial neural network with a set of 244 compounds and over one thousand calculated descriptors led to a prediction scheme used as a general in silico filter.[17] On the other hand, a decision tree-based approach was proposed using mainly three simple calculated physicochemical properties (*C*log *P*, CMR and pK_a) for classification of hERG channel blockers.[18] Upon analysis of a training set of 85 active hERG channel blocker plus 329 inactive compounds, a more elaborate scheme was set up in a 'veto' format of predictions from a combined pharmacophore fingerprint descriptor based model and topological 2D similarity nearest neighbors-type classification.[19] A hologram QSAR (HQSAR) model constructed from an initial 55 compound training set was claimed to be highly predictive.[20]

To make a quantitative in silico prediction on how a drug influences the hERG channel function is still a formidable task. A single pharmacophore-based approach may not be able to adequately describe structural determinants of multiple binding sites.[21,22] Furthermore,

* Corresponding author.

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.03.062

there is ample evidence to suggest that binding interactions of drug molecules at the pore cavity are dependent upon the conducting state of the channel.[23,24] Thus, a reliable model by chemometric approach requires a large data set of consistent experimental measurements. Accruement of our in-house data set containing diverse compounds prompted us to develop a general method using multiple models for the prediction of hERG channel blockers.

Our dataset consisted of 882 compounds with experimentally measured IC_{50} values for hERG channel inhibition. Standard whole-cell patch clamp electrophysiology method was used to record currents of hERG channels stably expressed in Chinese hamster ovary cells as described previously.[10] The number of compounds with zero, one, or two ionizable basic nitrogen atoms was 338, 499, and 45, respectively. Amide, aromatic and aniline nitrogen atoms were considered as non-basic. Activity data were reported as $-\log 10$ of IC_{50} (pIC_{50}), which ranged from -3.5 to 8.5.

We first attempted to model the dataset using the HQSAR method[25] implemented in the Sybyl QSAR program from Tripos. The HQSAR approach involves partial least squares (PLS) regression to correlate biological activities with molecular descriptors or holograms that encode a fixed length array containing counts of a priori defined substructural fragments. This method uses only 2D structure information, therefore avoiding the complication of conformational flexibility and structure alignments. As shown in Table 1, the best HQSAR model was obtained by correlating pIC_{50}s with a hologram containing 401 bins with six components in PLS regression. On the other hand, classification models using different cutoffs of hERG IC_{50} (100 nM, 1 or 10 uM) all led to poorer correlations in terms of cross-validated (leave one out) q^2 values. In contrast to the literature report, the HQSAR method failed to provide a predictive model for the hERG channel blockers from our data set.

To develop 3D pharmacophore models, the geometries of all molecules were initially generated by CONCORD and subsequently energy minimized using Tripos' software package SYBYL with the standard TRIPOS force field. Molecular interaction fields (MIFs) were then calculated using program GRID to determine energetically favorable interactions between the molecule and a probe group.[26] At each grid point, the interaction energy between the probe and the target molecule was calculated as the sum of Lennard-Jones potential, hydrogen bond, electrostatic interactions, and an entropic term.

Table 1. Performance of HQSAR models

Dependent variable	n[a]	q^2	r^2	SSE (cv)	SSE
pIC_{50}[a]	882	0.35	0.52	0.71	0.61
Binned at 100 nM	26+ (856)	<0	Na	Na	Na
Binned at 1 uM	232+ (650)	0.18	0.40	0.39	0.34
Binned at 10 uM	546+ (336)	0.25	0.43	0.43	0.37

[a] Number of 'inactive' compounds shown in parentheses.

MIFs obtained from GRID calculations were then transformed into alignment independent descriptors (GRIND).[27] The GRIND approach aimed to extract the information enclosed in the MIFs by compressing them into new types of variables whose values are independent of the spatial position of the molecule studied. Specifically, relevant regions were extracted from the MIFs using an optimization algorithm and a scoring function based on the intensity of the field at a node and the node-node distances between the selected nodes. From each field, many nodes (on the order of 100) were extracted to represent independent, favorable probe-ligand interaction regions. The products of the energy of interaction for each pair of nodes were also computed and assigned to a distance bin according to the node separation. By keeping only the highest product, autocorrelograms (from nodes by the same probe) and cross-correlograms (from nodes by different probes) were calculated for each distance bin, which allowed the original 3D information to be maintained as a line linking two specific MIF nodes. Therefore, these descriptors are ideally suited to represent pharmacodynamic properties in such a way that they are independent of alignment.

All the calculations were carried out using the program ALMOND (v3.2.0.) with four selected GRID probes: DRY (representing hydrophobic interactions), sp^2 carbonyl oxygen (representing H-bond acceptor), neutral flat amide NH (representing H-bond donor) and the TIP probe representing molecular shape descriptors. The shape descriptors were based on the local curvature of the molecular surface and were represented in a correlogram-like form.[28] Namely, autocorrelograms describe the distance between certain regions defined by the spatial extent of the molecule and cross-correlograms describe the distance between these regions and regions represented by other relevant interactions of the compounds. The grid spacing was set to 0.5 Å and the smoothing window was set to 0.8. The number of filtered nodes was set to 100 with 50% of relative weights. This yielded a total of ten groups of variables, including four autocorrelograms and six cross-correlograms. Correlations between the hERG channel inhibition and GRIND descriptors were analyzed using multivariate techniques such as principal component analysis (PCA) and PLS regression analysis.

In order to assess the importance of the presence of a charged nitrogen atom, the dataset was divided into two subsets. The first subset contained 338 compounds without any basic nitrogen atoms while the second set consisted of 544 molecules with one or two basic nitrogen atoms.

For the non-basic nitrogen subset, the 338 molecules were split into a training set (322 compounds) and a test set (16 molecules). To select the test set, the dataset was divided in 3 groups according to the activity value of pIC_{50} ($8 > 6.5$; $6.5 > 5$; $5 > 3.5$). The test set compounds were randomly chosen from each group to uniformly cover the activity range. The principal component analysis of the complete dataset was performed to analyze

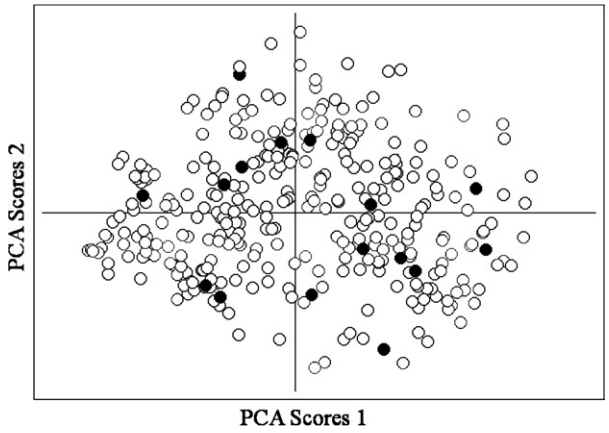

Figure 1. 2D plot of PCA showing the descriptor space of training set (open circle) and test set (filled dot) I (compounds without any basic nitrogen atoms).

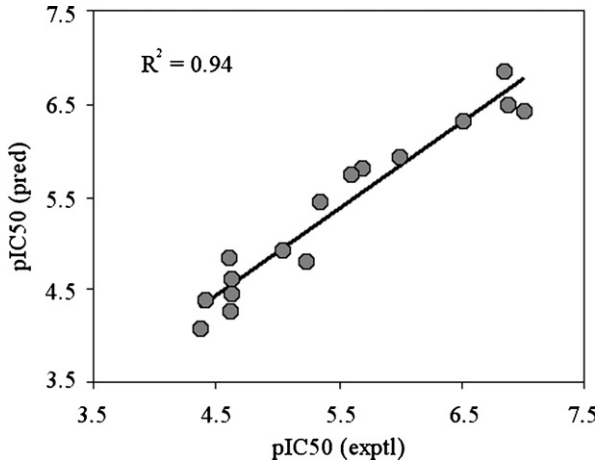

Figure 2. Calculated versus experimental pIC50 for 16 molecules of the test set using model for molecules with non-ionizable basic nitrogen.

structural variance of both the training and test sets. The PCA scores plot (Fig. 1) showed that there was no structural outlier present in the dataset and that training set and test set shared the same chemical space.

The PLS multivariate data analysis of the training set was carried out on the descriptors matrix to correlate the complete set of variables with the activity data. From a total of 710 variables, 559 active variables remained after filtering descriptors with no variability by Almond program. The PLS analysis resulted in a four latent variables (LVs) with an $r^2 = 0.76$. The cross validation of the model using leave-one-out (LOO) method yielded q^2 values of 0.72. As shown in Table 2, the GRIND descriptors 11-36, 44-49, 12-28, 13-42, 14-46, 24-46 and 34-45 were found to correlate with the inhibition activity in terms of high coefficients.

The predictive quality and robustness of the model were examined using 16 molecules of the test set defined previously. Figure 2 plots the predicted hERG activities of the test set molecules versus experimental measured values, showing a good agreement between the two.

For the ionizable basic nitrogen subset, all 544 molecules were protonated with a formal charge on the basic nitrogen. They were divided into a training set (518 compounds) and a test set (26 molecules). To select the test set, the dataset was divided in 3 groups according to the activity value of pIC_{50} ($8.5 > 7$; $7 > 5.5$; $5.5 > 3.9$). The test set compounds were randomly selected from each group to represent uniformly the activity span. The PCA of the datasets was performed to assess structural variance of both training and test sets. The PCA scores plot (Fig. 3) showed absence of any structural outlier in the dataset and that the training set and test set shared the same chemical space.

PLS analysis was performed on the training set to identify a correlation between the complete set of variables and the activity data. The ALMOND program kept 624 active variables out a total of 750 variables after filtering out descriptors with no variability. The 45 compounds with two basic nitrogen atoms were protonated on just one of the two basic centers. For these compounds, two models were built using one of the two isomers at a time. The isomer that produced the model with the best r^2 was chosen. The PLS analysis resulted in a model with three LVs and an $r^2 = 0.77$. The cross validation of the

Table 2. Salient GRIND descriptors in PLS models

GRIND descriptor	GRID MIFs	Subset I neutral		Subset II charged	
		Step	Å	Step	Å
11	Dry-dry	36	18	36	18
44	Tip-tip	49	24.5	58	29
12	Dry-Hbond donor	28	14	41	20.5
13	Dry-Hbond acceptor	42	21	40	20
14	Dry-tip	46	23	47	23.5
24	Hbond donor-tip	46	23	52	26
34	Hbond acceptor-tip	45	22.5	49	24.5

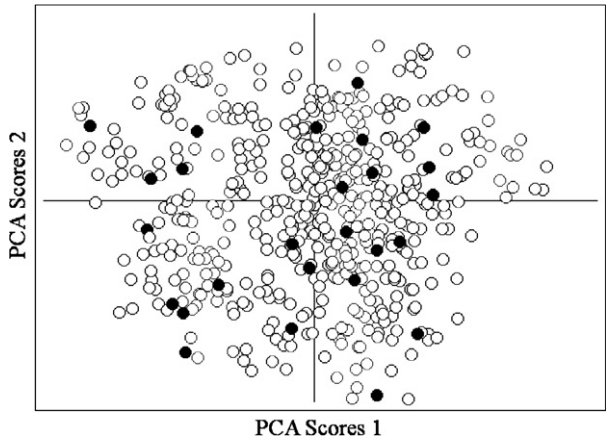

Figure 3. 2D plot of PCA showing the descriptor space of training set (open circle) and test set (filled dot) for subset II (molecules with protonable nitrogen atoms).

model by LOO method yielded q^2 values of 0.74. As shown in Table 2, the GRIND descriptors that had high coefficients in the PLS model were 11-36, 44-58, 12-41, 13-40, 14-47, 24-52, and 34-49.

For the model with charged basic nitrogen atoms, the predictive quality and robustness of the model were assessed using the test set of 26 molecules selected previously. Figure 4 plots the predicted HERG activities of the test set molecules versus experimental measured values.

Confirming the results of our previous CoMSiA model,[16] the charged nitrogen was found to be a relevant feature correlating the variance of the structural data to the activity data. PLS data analysis on the same training set but with all the basic nitrogen atoms in the neutral (non-protonated) form resulted in a PLS model of statistically lower quality than the one derived from the basic-nitrogen in a protonated form. This illustrated the importance of the positive charge to correlate the variance of the structural data to the activity value for this basic-nitrogen subset, in spite of the fact that the same kinds of descriptors were found to correlate with the activity data.

According to the two PLS models, all of the descriptors that correlated with the variance of the activity were very similar except for those involving an H-bond donor which corresponds to a basic protonated Nitrogen. For example, the distance between two hydrophobic MIFs is 18 Å, and the distance between the MIFs produced by a hydrophobic area and a H-bond acceptor group is ~21 Å in both the models.

Between the two models, there are differences in terms of descriptor distance between the edges of the molecule (tip-tip or GRIND descriptor 44) and the space between a hydrophobic MIF and a field generated by a H-bond donor group (dry-H-bond donor or GRIND descriptor 12). According to the PLS model built on the first subset (molecules devoid of any basic nitrogen atom), the optimal spacing between the fields generated by two edges of the molecule was ~25 Å, while the ideal size suggested by the second model was ~29 Å.

The distance between the MIFs produced by a hydrophobic and a H-bond-donor group was the most different between the two models. This also revealed an important pharmacophoric feature that defines the position of the groups that generate the dry-H-bond donor descriptors. For a molecule devoid of any ionizable basic-nitrogen group, the distance was ~14 Å, while the optimal distance between the same two MIFs was ~21 Å for molecules with a basic nitrogen atom. Figure 5 shows two inhibitors and the relative GRIND descriptors 12. For both compounds, high statistical relevance was always assigned to the field generated by the H-bond donor group close to the edge of the structure.

Our studies also suggested that H-bond acceptors play an important role for compounds to bind the hERG channel (GRIND descriptors 13 and 34 in both the pharmacophoric models, as shown in Fig. 6). The statistical relevance of the MIFs generated by the hydrophobic probe confirmed the assumptions that we made in our previous CoMSiA model[16] regarding the presence of a hydrophobic feature.

All the descriptors that play a role in the QSARs are almost the same except for those involving an H-bond donor, the basic protonated Nitrogen, perhaps suggesting the same binding pocket for both types of molecules. In the absence of a high quality X-ray structure of the hERG channel, one can speculate at best how the two classes of compounds interact with the protein. The PLS multivariate analysis was able to identify two pharmacophores that showed a high degree of similarity.

In summary, we obtained general 3D-QSAR models for hERG inhibitors, derived from GRIND pharmaco-

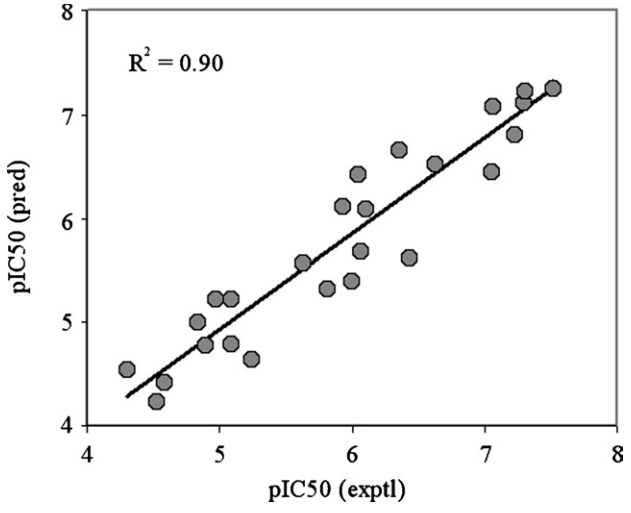

Figure 4. Calculated versus experimental pIC50 for 26 molecules of the test set using model for the charged basic nitrogen.

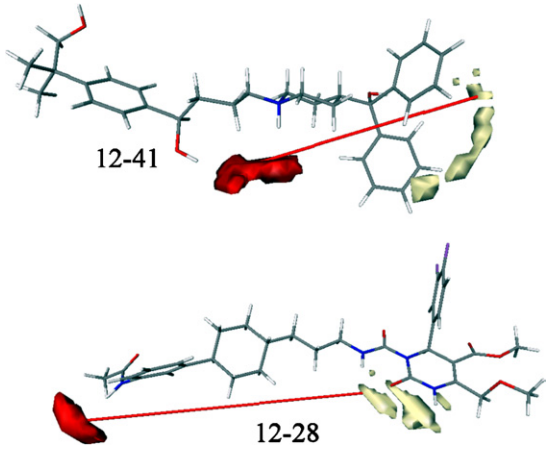

Figure 5. The pharmacophoric element (descriptor 12) from field generated by a H-bond donor group and hydrophobic area identified by the ALMOND models in mol11 (top) and mol316172 (bottom).

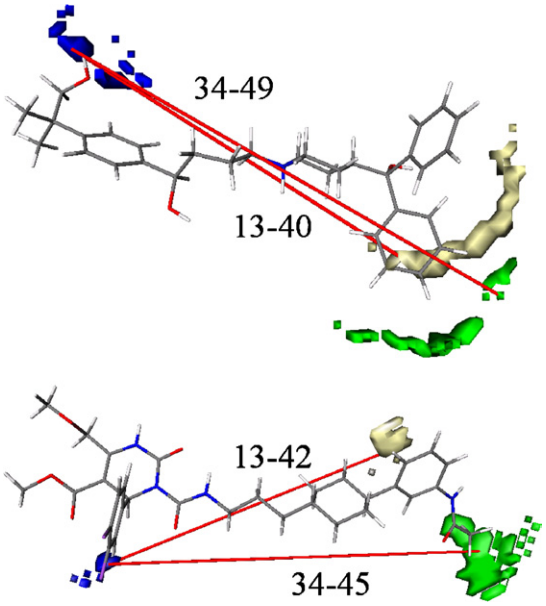

Figure 6. The GRIND descriptors for mol11 (top) and mol316172 (bottom) related to the presence of field generated by a H-bond acceptor group (blue), hydrophobic area (yellow), and TIP field (green).

phoric descriptors for datasets with and without a basic nitrogen.

Additional material

Compound	Experimental (pIC_{50}–6)	Calculated (pIC_{50}–6)
Sertindole	2.52	2.11
mol10[a]	2.21	1.87
mol3[a]	2.15	1.74
Dofetilide	2.00	1.94
mol2[a]	2.00	1.62
mol15[a]	1.96	0.33
Pimozide[a]	1.77	1.42
Sparfloxacin	1.74	1.56
mol13[a]	1.63	1.10
Haloperidole	1.49	0.86
MK499	1.47	1.61
mol8[a]	1.44	1.35
Cisapride	1.35	1.01
Terfenadine	1.25	0.83
mol1[a]	1.06	1.42
mol7[a]	0.88	1.09
mol6[a]	0.86	0.66
Verapamil	0.84	0.89
Ziprasidone	0.77	0.88
Thioridazine	0.72	0.91
Halofantrine[b]	0.71	0.54
mol14[a]	0.69	0.59
mol11[a]	0.34	0.18
Quinidine	0.26	0.20
Azimilide[b]	0.25	0.15
mol4	0.24	0.04
Cyamemazine	0.06	0.01
Chlorpromazine[b]	−0.17	−0.01
mol17[a]	−0.17	−0.76
mol19[a]	−0.29	−0.59
mol21[a]	−0.34	−0.79
Imipramine[b]	−0.53	−0.22
mol22[a]	−0.54	−0.96
mol18[a]	−0.66	−1.07
Mefloquine	−0.74	−0.33
Quetiapine	−0.76	−0.54
Terodiline	−0.76	0.11
Olanzapine	−0.78	−0.21
mol12[a]	−1.00	−1.04
mol23[a]	−1.00	−0.76
mol9[a]	−1.00	−0.97
mol20[a]	−1.20	−0.86
Fexofenadine	−1.36	−1.05
mol16[a]	−1.41	−1.27
Grepafloxacin	−1.70	−1.23
mol5[a]	−1.88	−1.62
Gatifloxacin	−2.11	−2.34
Moxifloxacin	−2.11	−2.03
Risperidone	−2.23	−1.89

[a]Ref. 4.
[b]Ref. 15.

Acknowledgments

We wish to thank Stephan Reiling for help with database manipulation.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2005.03.062.

References and notes

1. Sanguinetti, M. C.; Jiang, C.; Curran, M. E.; Keating, M. T. *Cell* **1995**, *81*, 299.
2. De Ponti, F.; Poluzzi, E.; Cavalli, A.; Recanatini, M.; Montanaro, N. *Drug Safety* **2002**, *25*, 263.
3. Mitcheson, J. S.; Perry, M. D. *Curr. Opin. Drug Discov. Dev.* **2003**, *6*, 667.
4. Pearlstein, R.; Vaz, R.; Rampe, D. *J. Med. Chem.* **2003**, *46*, 2017.
5. Netzer, R.; Bischoff, U.; Ebneth, A. *Curr. Opin. Drug Discov. Dev.* **2003**, *6*, 462.
6. Tang, W.; Kang, J.; Wu, X.; Rampe, D.; Wang, L.; Shen, H.; Li, Z.; Dunnington, D.; Garyantes, T. *J. Biomol. Screen.* **2001**, *6*, 325.
7. Cheng, C. S.; Alderman, D.; Kwash, J.; Dessaint, J.; Patel, R.; Lescoe, M. K.; Kinrade, M. B.; Yu, W. *Drug. Dev. Ind. Pharm.* **2002**, *28*, 177.
8. Finlayson, K.; Turnbull, L.; January, C. T.; Sharkey, J.; Kelly, J. S. *Eur. J. Pharmacol.* **2001**, *430*, 147.

9. Chiu, P. J.; Marcoe, K. F.; Bounds, S. E.; Lin, C. H.; Feng, J. J.; Lin, A.; Cheng, F. C.; Crumb, W. J.; Mitchell, R. *J. Pharmacol. Sci.* **2004**, *95*, 311.
10. Kang, J.; Wang, L.; Cai, F.; Rampe, D. *Eur. J. Pharmacol.* **2000**, *392*, 137.
11. Zou, A.; Curran, M. E.; Keating, M. T.; Sanguinetti, M. C. *Am. J. Physiol.* **1997**, *272*, 1309.
12. Kiss, L.; Bennett, P. B.; Uebele, V. N.; Koblan, K. S.; Kane, S. A.; Neagle, B.; Schroeder, K. *Assay Drug Dev. Technol.* **2003**, *1*, 127.
13. Recanatini, M.; Poluzzi, E.; Masetti, M.; Cavalli, A.; De Ponti, F. *Med. Res. Rev.*Published online: 16 Aug 2004.
14. Ekins, S.; Crumb, W. J.; Sarazan, R. D.; Wikel, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427.
15. Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. *J. Med. Chem.* **2002**, *45*, 3844.
16. Pearlstein, R. A.; Vaz, R. J.; Kang, J.; Chen, X. L.; Preobrazhenskaya, M.; Shchekotikhin, A. E.; Korolev, A. M.; Lysenkova, L. N.; Miroshnikova, O. V.; Hendrix, J.; Rampe, D. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1829.
17. Roche, O.; Trube, G.; Zuegge, J.; Pflimlin, P.; Alanine, A.; Schneider, G. *ChemBioChem.* **2002**, *3*, 455.
18. Buyck, C.; Tollenaere, J.; Engels, M.; De Clerck, F. In *The 14th European Symposium on Quantitative Structure–Activity Relationships*: Bournemouth, UK, 2002.
19. Aronov, A. M.; Goldman, B. B. *Bioorg. Med. Chem.* **2004**, *12*, 2307.
20. Keseru, G. M. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773.
21. Korolkova, Y. V.; Bocharov, E. V.; Angelo, K.; Maslennikov, I. V.; Grinenko, O. V.; Lipkin, A. V.; Nosyreva, E. D.; Pluzhnikov, K. A.; Olesen, S. P.; Arseniev, A. S.; Grishin, E. V. *J. Biol. Chem.* **2002**, *277*, 43104.
22. Pardo-Lopez, L.; Zhang, M.; Liu, J.; Jiang, M.; Possani, L. D.; Tseng, G. N. *J. Biol. Chem.* **2002**, *277*, 16403.
23. Kamiya, K.; Mitcheson, J. S.; Yasui, K.; Kodama, I.; Sanguinetti, M. C. *Mol. Pharmacol.* **2001**, *60*, 244.
24. Witchel, H. J.; Dempsey, C. E.; Sessions, R. B.; Perry, M.; Milnes, J. T.; Hancox, J. C.; Mitcheson, J. S. *Mol. Pharmacol.* **2004**, *66*, 1201.
25. Heritage, T. W.; Lowis, D. R. In *Rational Drug Design: Novel Methodology and Practical Applications*; Parrill, A. L., Reddy, M. R., Eds.; American Chemical Society: Washington, DC, 1999.
26. Wade, R. C.; Goodford, P. J. *Prog. Clin. Biol. Res.* **1989**, *289*, 433.
27. Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43*, 3233.
28. Fontaine, F.; Pastor, M.; Sanz, F. *J. Med. Chem.* **2004**, *47*, 2805.