molecular pharmaceutics

articles

hERG Classification Model Based on a Combination of Support Vector Machine Method and GRIND Descriptors

Qiyuan Li,[†] Flemming Steen Jørgensen,[‡] Tudor Oprea,[§] Søren Brunak,[†] and Olivier Taboureau*,[†,‡]

Center for Biological Sequence Analysis, Biocentrum-DTU, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark, Department of Medicinal Chemistry, The Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark, and Division of Biocomputing, Department of Biochemistry and Molecular Biology, University of New Mexico School of Medicine, MSC11 6145, Albuquerque, New Mexico 87131

Received September 13, 2007; Revised Manuscript Received November 1, 2007; Accepted November 12, 2007

Abstract: The human Ether-a-go-go Related Gene (hERG) potassium channel is one of the major critical factors associated with QT interval prolongation and development of arrhythmia called Torsades de Pointes (TdP). It has become a growing concern of both regulatory agencies and pharmaceutical industries who invest substantial effort in the assessment of cardiac toxicity of drugs. The development of *in silico* tools to filter out potential hERG channel inhibitors in early stages of the drug discovery process is of considerable interest. Here, we describe binary classification models based on a large and diverse library of 495 compounds. The models combine pharmacophore-based GRIND descriptors with a support vector machine (SVM) classifier in order to discriminate between hERG blockers and nonblockers. Our models were applied at different thresholds from 1 to 40 μm and achieved an overall accuracy up to 94% with a Matthews coefficient correlation (MCC) of 0.86 (*F*-measure of 0.90 for blockers and 0.95 for nonblockers). The model at a 40 μm threshold showed the best performance and was validated internally (MCC of 0.40 and *F*-measure of 0.57 for blockers and 0.81 for nonblockers, using a leave-one-out cross-validation). On an external set of 66 compounds, 72% of the set was correctly predicted (*F*-measure of 0.86 and 0.34 for blockers and nonblockers, respectively). Finally, the model was also tested on a large set of hERG bioassay data recently made publicly available on PubChem (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=376) to achieve about 73% accuracy (*F*-measure of 0.30 and 0.83 for blockers and nonblockers, respectively). Even if there is still some limitation in the assessment of hERG blockers, the performance of our model shows an improvement between 10% and 20% in the prediction of blockers compared to other methods, which can be useful in the filtering of potential hERG channel inhibitors.

Keywords: Human Ether-a-go-go Related Gene; support vector machine; GRIND descriptors; classification model

Introduction

The development of acquired long QT syndrome and Torsades de Pointes (TdP), both recognized as electrocardiac symptoms of cardiotoxicity, are mediated in part by the blockage of the voltage-dependent potassium ion channel encoded by the human Ether-a-go-go Related Gene (hERG).[1–3] A large number of compounds covering a broad spectrum of structures, including fluoroquinolone antibiotics, anti-

* To whom correspondence should be addressed. Mailing address: Center for Biological Sequence Analysis, Biocentrum-DTU, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark. Tel: +45 45 25 61 67. Fax: +45 35 30 61 62. E-mail: otab@cbs.dtu.dk.

[†] Technical University of Denmark.

[‡] University of Copenhagen.

[§] University of New Mexico School of Medicine.

[1] Sanguinetti, M. C.; Jiang, C.; Curran, M. E.; Keating, M. T. A mechanistic link between an inherited and an acquired cardiac arrhythmia: HERG encoded the Ikr potassium channel. *Cell. Physiol. Biochem.* **1995**, *81*, 299–307.

[2] Brown, A. M. Drugs, hERG and sudden death. *Cell Calcium.* **2004**, *35*, 543–547.

10.1021/mp700124e CCC: $37.00 © XXXX American Chemical Society
Published on Web 01/16/2008

psychotic agents, antihistamines, as well as Class-III antiarrhytmics, are known as hERG channel blockers.[4,5] Nowadays, the assessment of the hERG blocking potential of novel chemical structures is monitored by all major pharmaceutical companies during each stage of the drug discovery process,[6] though not all compounds known to block hERG cause QT prolongation over 10–20 ms and, consequently, TdP. However, reliable experimental determination of hERG channel inhibition by a drug requires whole cell patch clamp electrophysiology studies, which are time-consuming and expensive.[7] That is why over the past few years considerable effort has been devoted to the understanding of the mechanism and the structural requirements for hERG blockage. The hERG channel is formed by four identical α-subunits, each containing six α-helical transmembrane domains, S1−S6. Segments S1−S4 form the voltage sensor part where the movement of the gating is driven by the positively charged Lys and Arg in the S4 helix whereas segments S5−S6 form the central channel cavity where blockers bind.[8] Since no crystal structure of the hERG channel is yet available, homology models based on bacterial crystal structures like the K+ channel from *Streptomyces lividans* (KcsA),[9] the bacterial K+ channel (KvAP),[10] the K+ channel from *Methanobacterium thermoautotrophicum* (MthK),[11] and recently a mammalian Kv1.2 K+ channel[12] were performed. They provided a general understanding of the nature of the interactions involved like the $\pi-\pi$ interaction between a positively charged nitrogen of some drugs and Tyr652 and also the hydrophobic attraction with Phe656.[13] Site-directed mutagenesis studies have also allowed the identification of key residues responsible for the high affinity interaction of hERG blockers. The mutations of Tyr652 and Phe656 to Ala, both located in the S6 transmembrane helix, decrease the affinity of all potent known blockers.[14] Additional residues like Val625 and Gly648 seem to more specifically affect MK-499, clofilium, and ibutilide.[15] With the gain of information about hERG inhibition for a large set of compounds, the development of computational models to predict hERG activity has been intensively investigated. Binary classification models based on neural networks[16,17] 2D-QSAR[18–21] and 3D-QSAR[22–24] associated with different sets of descriptors (fragment based, topological and physicochemical parameters, and molecular interaction field descriptors) were realized in order to discriminate hERG channel blockers. Recently, 3D-QSAR models derived from GRIND pharmacophoric descriptors

[3] Perlstein, R. A.; Vaz, R. J.; Rampe, D. Understanding the structure-activity relationship of the human ether-a-go-go-related gene cardiac K+ channel. A model for bad behaviour. *J. Med. Chem.* **2003**, *46*, 2017–2022.

[4] De Ponti, F.; Poluzzi, E.; Cavalli, A.; Recanatini, M.; Montanaro, N. Safety of non-antiarrhytmic drugs that prolong the QT interval or induce torsades de pointes: an overview. *Drug Safety* **2002**, *25*, 263–286.

[5] De Ponti, F.; Poluzzi, E.; Montanaro, N. QT-interval prolongation by non cardiac drugs: lessons to be learned from recent experience. *Eur. J. Clin. Pharmacol.* **2000**, *56*, 1–18.

[6] Crumb, W.; Cavero, I. QT interval prolongation by non cardiovascular drugs: issues and solutions for novel drug development. *Pharm. Sci. Technol. Today* **1999**, *2*, 270–280.

[7] Cavero, I.; Crumb, W. Native and cloned ion channels from human heart: laboratory models for evaluating the cardiac safety of new drugs. *Eur. Heart J.* **2001**, *3*, 53–63.

[8] Sanguinetti, M. C.; Tristani-Firouzi, M. HERG potassium channels and cardiac arrhythmia. *Nature* **2006**, *440*, 463–469.

[9] Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The structure of the potassium channel molecular basis of K+ coduction and selectivity. *Science* **1998**, *280*, 69–77.

[10] Jiang, Y.; Lee, A.; Chen, J.; Cadene, M.; Chait, B. T.; MacKinnon, R. Crystal structure and mechanism of a calcium-gated potassium channel. *Nature* **2002**, *417*, 515–522.

[11] Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R. X-ray structure of a voltage-dependent K+ channel. *Nature* **2003**, *423*, 33–41.

[12] Long, S. B.; Campbell, E. B.; MacKinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science* **2005**, *309*, 897–903.

[13] Fernandez, D.; Ghanta, A.; Kauffman, G. W.; Sanguinetti, M. C. Physicochemical features of the hERG channel drug binding site. *J. Biol. Chem.* **2004**, *279*, 10120–10127.

[14] Mitcheson, J. S.; Chen, J.; Lin, M.; Culberson, C.; Sanguinetti, M. C. A structural basis for drug-induced long QT syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12329–12333.

[15] Perry, M.; De Groot, M. J.; Heliwell, R.; Leishman, D.; Tristani-Firouzi, M.; Sanguinetti, M. C.; Mitcheson, J. S. Structural determinants of HERG channel block by clofilium and ibutilide. *J. Mol. Pharmacol.* **2004**, *6*, 240–249.

[16] Aronov, A. M.; Goldman, B. B. A model for identifying HERG K+ channel blockers. *Bioorg. Med. Chem.* **2004**, *12*, 2307–2315.

[17] Tobita, M.; Nishikawa, T.; Nagashima, R. A discriminant model constructed by the support vector machine method for hERG potassium channel inhibitors. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2886–2890.

[18] Keresu, G. M. Prediction of hERG potassium channel affinity by traditional and hologram QSAR methods. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2773–2775.

[19] Bains, W.; Basman, A.; White, C. HERG binding specificity and binding site structure: evidence from a fragment-based evolutionary computing SAR study. *Prog. Biophys. Mol. Biol.* **2004**, *86*, 205–233.

[20] Song, M.; Clark, M. Development and evaluation of an in silico model for hERG binding. *J. Chem. Inf. Model.* **2006**, *46*, 392–400.

[21] Seierstad, M.; Agrafiotis, D. K. A QSAR model of hERG binding using a large, diverse, and internally consistent training set. *Chem. Biol. Des.* **2006**, *67*, 284–296.

[22] Ekins, S.; Crumb, W. J.; Sarazan, R. D.; Wikel, J. H.; Wrighton, S. A. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a go-go-related gene potassium channel. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427–434.

[23] Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. Towards a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of hERG K+ channel blockers. *J. Med. Chem.* **2002**, *45*, 3844–3853.

[24] Pearlstein, R. A.; Vaz, R. J.; Kang, J.; Chen, X. L.; Preobrazhenskaya, M.; Shchekotikhin, A. E.; Korolev, A. M. Characterization of hERG potassium inhibition using CoMSIA 3D QSAR and homology modeling approaches. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1829–1835.

were reported to be predictive.[25] The advantage of this approach is that is allows for the representation of pharmacophoric properties in such a way that they are independent of alignment. Usually, it is quite common to use linear regression, namely partial least-squares (PLS), to correlate the GRIND descriptors to the experimental measurements (IC_{50}, K_i) of the data set studied. However, this 3D-QSAR method is quite sensitive to the activity considered, and the introduction of uncertainty in the data set can generate some drastic errors. A classification method is more flexible and can minimize the uncertainty of the activity. Moreover, it is suitable for filtering hERG blockers in a qualitative way from chemical libraries and virtual chemical databases before using more accurate methods.

With the aim of developing a classification model to discriminate hERG blockers from nonblockers, we report here the combination of the GRIND descriptors with a support vector machine (SVM) method for classification of a large and structurally diverse set of compounds with known hERG activity. To do so, we have collected 495 compounds with hERG information from the literature, and all of the molecules were docked in a homology model of the homotetrameric pore domain (S5−S6) of the hERG channel based on the crystal structure of the bacterial potassium channel KvAP. This step provides reasonable conformations of the molecules in the data set related to the hydrophobic environment of the pore domain. Pharmacophoric GRIND descriptors computed for each compound were then associated with a SVM classifier at different thresholds of hERG inhibitor activity to define the most reliable model. The models were then applied to two external sets in the order to evaluate their performance. To our knowledge, it is the first time the combination of GRIND descriptors and a SVM method has been implemented with good classification accuracy at the result and validated on a large external set recently made available on the PubChem bioassay database.

Material and Methods

Data Sets. The majority of our data set (322 compounds) comes from Aronov and co-workers.[16] We updated the collection with a subset of 173 compounds which were tested experimentally in recent publications[17–38] to have in total a training set of 495 structurally diverse compounds. We have primarily recovered compounds determined by the patch clamp hERG current inhibition assay using the mammalian cell lines, HEK, CHO, COS, or neuroblastoma cells. However, a few drugs with a measurement from nonmammalian cell lines, XO (*Xenopus laevis* oocytes), were also included when mammalian cell data were not available. The ideal training set for any model needs to be preferably large, diverse, and consistent, but since the goal was to create a qualitative virtual screen as opposed to a quantitative model for predicting hERG inhibition, the classification approach is expected to be less sensitive to small differences from one study to another.[16]

We have tested binary classification models at thresholds of 1, 5, 10, 20, 30, and 40 μm, respectively, to define the most suitable cutoff to discriminate between hERG blockers and nonblockers with this data set. A cutoff up to 50 μm would not typically be considered clinically relevant. An additional test set, 66 compounds from the WOMBAT-PK

(25) Cianchetta, G.; Li, Y.; Kang, J.; Rampe, D.; Fravolini, A.; Cruciani, G.; Vaz, R. J. Predictive models for hERG potassium channel blockers. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3637–3642.

(26) Fenichel, R. http://fenichel.net/pages/Professional/subpages/QT/Tables/pbydrug.htm, 2006.

(27) Yuill, K. H.; Borg, J. J.; Ridley, J. M.; Milnes, J. T.; Witchel, H. J.; Paul, A. A.; Kozlowski, R. Z.; Hancox, J. C. Potent inhibition of human cardiac potassium (HERG) channels by the anti-estrogen agent clomiphene-without QT interval prolongation. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 556–561.

(28) Waldegger, S.; Niemeyer, G.; Morike, K.; Wagner, C. A.; Suessbrich, H.; Busch, A. E.; Lang, F.; Eichelbaum, M. Effect of verapamil enantiomers and metabolites on cardiac K+ channels expressed in Xenopus oocytes. *Cell Physiol. Biochem.* **1999**, *9*, 81–89.

(29) Katchman, A. N.; McGroary, K. A.; Kilborn, M. J.; Kornick, C. A.; Manfredi, P. L.; Woosley, R. L.; Ebert, S. N. Influence of opioid agonists on cardiac human ether-a-go-go related gene K(+) currents. *J. Pharmacol. Exp.Ther.* **2002**, *303*, 688–694.

(30) Milnes, J. T.; Crociani, O.; Arcangeli, A.; Hancox, J. C.; Witchel, H. J. Blockade of HERG potassium current by fluvoxamine: incomplete attenuation by S6 mutations at F656 or Y652. *Br. J. Pharmacol.* **2003**, *139*, 887–898.

(31) Claassen, S.; Zunkler, B. J. Comparison of the effects of metoclopramide and domperidone on HERG channels. *Pharmacology* **2005**, *74*, 31–36.

(32) Wible, B. A.; Hawryluk, P.; Ficker, E.; Kuyshev, Y. A.; Kirch, G.; Brown, A. M. HERG-Lite: A novel comprehensive high-throughput screen for drug-induced hERG risk. *J. Pharmacol. Toxicol. Methods* **2005**, *52*, 136–145.

(33) Redfern, W. S.; Carlsson, L.; David, A. S.; Lynch, W. G.; MacKenzie, I.; Palethorpe, S.; Siegl, P. K.; Strang, I.; Sullivan, A. T.; Wallis, R.; Camm, A. J.; Hammond, T. G. Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs: evidence for a provisional safety margin in drug development. *Cardiovasc. Res.* **2003**, *58*, 32–45.

(34) Chapman, H.; Pasternack, M. The action of the novel gastrointestinal prokinetic prucalopride on the HERG K(+) channel and the common T897 polymorph. *Eur. J. Pharmacol.* **2007**, *554*, 98–105.

(35) Kang, J.; Chen, X. L.; Wang, H.; Ji, J.; Reynolds, W.; Lim, S.; Hendrix, J.; Rampe, D. Cardiac ion channel effects of tolterodine. *J. Pharmacol. Exp. Ther.* **2004**, *308*, 935–940.

(36) Ekins, S.; Crumb, W. J.; Sarazan, R. D.; Wikel, J. H.; Wrighton, S. A. Three-dimensional quantitative structure-activity relationship for inhibition of human ether-a-go-go related gene potassium channel. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427–434.

(37) Caballero, R.; Moreno, I.; Gonzalez, T.; Arias, C.; Valenzuela, C.; Delpon, E.; Tamargo, J. Spironolactone and its main metabolite, canrenoic acid, block human ether-a-go-go-related gene channels. *Circulation* **2003**, *107*, 889–895.

(38) Yao, X.; McIntyre, M. S.; Lang, D. G.; Song, I. H.; Becherer, J. D.; Hashim, M. A. Propanolol inhibits the human ether-a-go-go related gene potassium channels. *Eur. J. Pharmacol.* **2005**, *519*, 208–211.

database[39] with hERG activity, was used to validate the performance of the models. These compounds come from other sources compared to those used in the training set with experimental binding activities evaluated in mammalian or nonmammalian cell lines and expressed by an IC_{50}, K_i, or even a percentage of current inhibition.[40–44] Recently, a large set of compounds with hERG activity became available on the PubChem bioassay database (http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?aid=376).[45] This data set contains 1948 compounds of which 248 are described as active and 1700 as nonactive and will be used to evaluate the accuracy of our models.

hERG Model and Docking. A homology model of the homotetrameric pore domain of hERG potassium channel, based on the available crystal structure of the bacterial KvaP channel in an open state proposed by Österberg and Åqvist,[46] was used to dock all the molecules in the data set. The docking of the 561 molecules (495 from the training set and 66 from the external set) was performed automatically using the MOE software.[47] Only ligands were flexible in this process, whereas the entire inner cavity of hERG was kept rigid. The docking module in MOE uses an affinity dG scoring function to assess candidate poses. This function estimates the enthalpy contribution to the free energy of binding using a linear model

$$G = C_{hb}f_{hb} + C_{ion}f_{ion} + C_{mlig}f_{mlig} + C_{hh}f_{hh} + C_{hp}f_{hp} + C_{aa}f_{aa}$$

where the f terms fractionally count atomic contacts of specific types and the C's are coefficients that weight the contributions of each term to the affinity. The individual terms are hb for interactions between hydrogen bond donor–acceptors pairs, ion for ionic interactions, mlig for metal ligation, hh for hydrophobic interactions, hp for hydrophobic−polar interactions, and aa for two atomic interactions. As our main goal is to use the docking process to obtain a suitable 3D binding conformation of each molecule rapidly, the number of docking poses was set to 10 and the pose with the best score was considered for further investigations with GRIND. It should be noticed that compounds with a basic nitrogen were protonated before the docking process as previous studies demonstrated the importance of a cation−π interaction between a positively charged nitrogen of a drug and the π-electrons of Tyr652.[48]

GRIND Descriptors. Based on the 3D conformation of the compounds, the molecular interaction fields (MIFs) were then calculated using the program GRID to determine energetically favorable interactions between each molecule and a probe group.[49] Four GRID probes were selected: DRY representing hydrophobic interactions, O sp2 carbonyl oxygen (representing H-bond acceptor), NH neutral flat amide (representing H-bond donor), and the TIP probe representing molecular shape descriptors. The GRIND approach extracts the information enclosed in the MIFs and encodes it into new variables whose values are independent of the spatial position of the molecule studied. These variables correspond to a product of the favorable energy of interaction region assigned to a distance between these regions. The highest product can result from two regions defined by the same probe (autocorrelogram) or by two different probes (cross-correlogram). Therefore, this approach

(39) Olah, M.; Rad, R.; Ostopovici, L.; Bora, A.; Hadaruga, N.; Hadaruga, D.; Ramona, R.; Fulias, A.; Mracec, M.; Oprea, T. I. WOMBAT and WOMBAT-PK: Bioactivity Databases for Lead and Drug Discovery. In *Chemical Biology: From Small Molecules to Systems Biology and Drug Design*; Schreiber, S. L., Kapoor, T. M., Wess, G., Eds.; Wiley-VCH: Weinheim, 2007; pp 760–786. The WOMBAT-PK database is available from Sunset Molecular Discovery LLC, Santa Fe, NM, http://www.sunsetmolecular.com, 2006.

(40) Rowley, M.; Hallett, D. J.; Goodacre, S.; Crawforth, J.; Sparey, T. J.; Patel, S.; Marwood, R.; Thomas, S.; Hitzel, L.; O'Connor, D.; Szeto, N.; Castro, J. L.; Hutson, P. H.; MacLeod, A. M. 3-(4-Fluoropiperidin-3-yl)-2-phenylindoles as High Affinity Selective and Orally Bioavailable h5-HT2a receptor Antagonists. *J. Med. Chem.* **2001**, *44*, 1603–1614.

(41) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beshore, D. C.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hartman, G. D.; Heimbrook, D. C.; Hommick, C. F.; Huff, J. R.; Kassahun, K.; Koblan, K. S.; Kohl, N. E.; Lobell, R. B., Jr.; Miller, P. A.; Omer, C. A.; Rodrigues, A. D.; Walsh, E. S.; William, T. M. Design and Biological activity of (S)-4–5-{[1- (3-Chlorobenzyl)-2-oxopyrrolidin-3-ylamino]methyl)imidazol-1-ylmethyl) benzonitrile, a 3-aminopyrrolidinone Farnesyltransferase Inhibitor with Excellent Cell potency. *J. Med. Chem.* **2001**, *44*, 2933–2949.

(42) Bell, I. M.; Gallicchio, S. N.; Abrams, M.; Beese, L. S.; Beshore, D. C.; Bhimnathwala, H.; Boqusky, M. J.; Buser, C. A.; Culberson, J. C.; Davide, J.; Ellis-Hutchings, M.; Fernandes, C.; Gibbs, J. B.; Graham, S. L.; Hamilton, K. A.; Hartman, G. D.; Heimbrook, D. C.; Homnick, C. F.; Huber, H. E.; Huff, J. R.; Kassahun, K.; Koblan, K. S.; Kohl, N. E.; Lobell, R. B., Jr.; Robinson, R.; Rodrigues, A. D.; Taylor, J. S.; Walsh, E. S.; Williams, T. M.; Zartman, C. B. 3-aminopyrrolidinone farnesyltransferase inhibitors: design of macrocyclic compounds with improved pharmacokinetics and excellent cell potency. *J. Med. Chem.* **2002**, *45*, 2388–2409.

(43) Peukert, S.; Brendel, J.; Pirard, B.; Brüggemamm, A.; Below, P.; Kleemann, H. W.; Hemmerle, H.; Schmidt, W. Identification, synthesis and activity of novel blockers of the voltage-gated potassium channel kv1.5. *J. Med. Chem.* **2003**, *46*, 486–498.

(44) Blum, C. A.; Zheng, X.; De Lombaert, S. Design, Synthesis, and Biological Evaluation of substituted 2-Cyclohexyl-4-phenyl-1H-imidazoles: Potent and selective Neuropeptide Y Y5-receptor Antagonists. *J. Med. Chem.* **2004**, *47*, 2318–2325.

(45) Brown, A. M. HERG block, QT liability and sudden cardiac death. *Novartis Found. Symp.* **2005**, *266*, 118–131, discussion 131–115, 155–118.

(46) Österberg, F.; Åqvist, J. Exploring blocker binding to a homology model of the open hERG K+ channel using docking and molecular dynamics methods. *FEBS Lett.* **2005**, *579*, 2939–2944.

(47) MOE: Molecular Operating Environment. http://www.chemcomp.com.

(48) Farid, R.; Day, T.; Friesner, R. A.; Pearlstein, R. A. New insights about HERG blockage obtained from protein modeling, potential energy mapping, and docking studies. *Bioorg. Med. Chem.* **2006**, *14*, 3160–3173.

(49) Wade, R. C.; Goodford, P. J. The role of hydrogen-bonds in drug binding. *Prog. Clin. Biol. Res.* **1989**, *289*, 433–444.

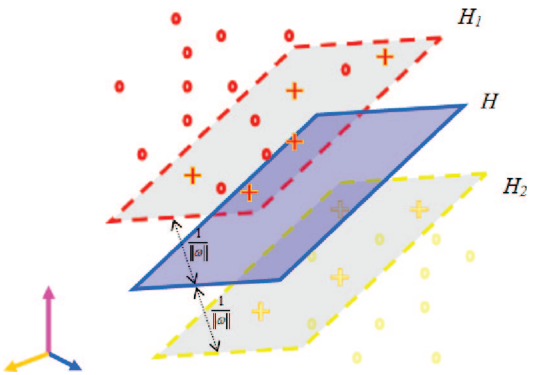


Figure 1. Principle behind SVM classification: H, classifying hyperplane; H_1 and H_2, auxiliary classifying plane; red/yellow ○, sample points from two classes in the space; red/yellow +, support vectors from two classes, defined as sample points in H_1 and H_2.

is ideally suitable to represent pharmacophoric properties independently of alignment. More information about the technique can be found in the work by Pastor et al.[50]

All of the calculations were carried out using the program ALMOND (v3.3.0). The grid spacing was set to 0.5 Å with the four probes indicated previously. This results in a total of 10 groups of variables, representing 870 variables, including four autocorrelograms and six cross-correlograms.

Classification with the SVM Method. From the 870 variables, 618 active variables were extracted and implemented in a SVM classification method. SVM is a statistical learning method developed by Vapnik et al. during the 1990s.[51] SVM provides a classifier that minimizes the so-called "structural risk" in prediction.[52] The principle of SVM classifier can be described as follows:

The points from two classes are classified in the sample space by a hyperplane H, and the optimized classifier plane is defined as the one that maximizes the margin between the two classes, which is measured by the distances between plane H and the planes cutting the nearest sample points on both sides of H, namely, H_1 and H_2. In particular, the sample points located exactly on planes H_1 and H_2 are defined as support vectors (Figure 1). In our study, we used both linear and nonlinear SVM classifiers associated with a radial bias function (RBF) implemented in the WEKA package.[53] After optimization, the complexity parameter was set to 1.0, $\gamma = 0.01$, a tolerance parameter of 0.001, and an exponent for the polynomial kernel of 1.0 for the linear method and 2.0 for a nonlinear SVM.[54,55] For classification, overall accuracy, sensitivity (prediction of correct hERG blockers among hERG blockers), and specificity (prediction of correct hERG nonblockers among hERG nonblockers) was considered. In addition, the Matthews coefficient correlation (MCC) was also determined as well as the F-measures to take into account the relative costs of false positives and false negatives in the model. All of these statistic values are between 0 and 1 with no correlation at 0 and perfect correlation at 1.

Results and Discussion

Evaluation of the Data Set. Since the data set used in this project came from different sources, we examined the correlation between the data and assessed the variations caused by different experimental conditions and instrumental errors. Such variations, if too large, can have a significant negative effect on a predictive model developed from the data. Figure 2 shows how the majority of the published data sets of hERG activity values correlate with each other. As a result, the Pearson correlation coefficients between different pairs of data sets vary from 0.57 to 0.99. In general, the data from one source to another are consistent within +/− 1 log unit if we consider only mammalian cells. The variation can be more significant if we compare compounds with activity measured in mammalian cell and nonmammalian cells. For example, pIC_{50} values of 6.76 and 4 were reported for loratidine in HEK[56] and XO[57] cells, respectively. Fortunately, only a few compounds tested solely in XO cells are present in our data set (only the carvedilol, epinastine, and verapamil metabolites). Thus, the qualitative approach of our method is expected to be less sensitive to these variations compared to a QSAR model. It is worthwhile to keep in mind that the predictions are based on prior knowledge and are useful as long as they are within the applicable domain.[58]

Another important factor to take into account is the diversity of the data set. Too many similar structures in the data set can affect the training of the model and may result in an overestimation of the predictive performance. In order

(50) Pastor, M.; Cruiciani, G.; McLay, I.; Pickett, S.; Clementi, S. GRid-Independent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.

(51) Cortes, C.; Vapnik, V. Support-Vector Networks. *Machine Learning* **1995**, *20*, 273–297.

(52) Noble, W. S. What is a support vector machine. *Nat. Biotechnol.* **2006**, *24*, 1565–1567.

(53) Witten, I. H.; Frank, E. *Data Mining: Practical machine learning tools with Java implementation*; Morgan Kaufmann: San Francisco, 2000.

(54) Platt, J. *Fast training of support vector machines using sequential minimal optimization. Advances in Kernel methods - Support Vector Learning*; Schoelkopf, B., Burges, C., Smola A., Eds.; MIT Press: Cambridge, MA, 1998.

(55) Keerthi, S. S.; Shevade, S. K.; Bhattacharyya, C.; Murthy, K. R. K. Improvements to Platt's SMO algorithm for SVM classifier design. *Neural Comput.* **2001**, *13* (3), 637–649.

(56) Crumb, W. J., Jr. Loratidine blockade of K^+ channels in human heart: comparison with terfenadine under physiological conditions. *J. Pharmacol. Exp. Ther.* **2000**, *292* (1), 261–264.

(57) Taglialatela, M.; Pannaccione, A.; Castaldo, P.; Giorgio, G.; Zhou, Z.; January, C. T.; Genovese, A.; Marone, G.; Annunziato, L. Molecular basis for the lack of HERG K+ channel block-related cardiotoxicity by the H_1 receptor-blocker cetirizine compared with other second-generation antihistamines. *Mol. Pharmacol.* **1998**, *54*, 113–121.

(58) Golbraikh, A.; Tropsha, A. Beware of q2! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.

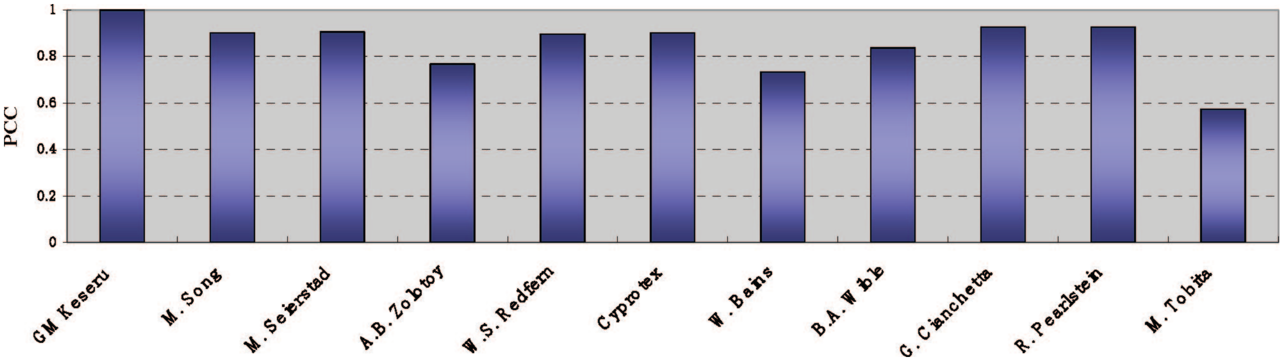


Figure 2. Histogram of Pearson's correlation between Fenichel's data set and each of the remaining ones.

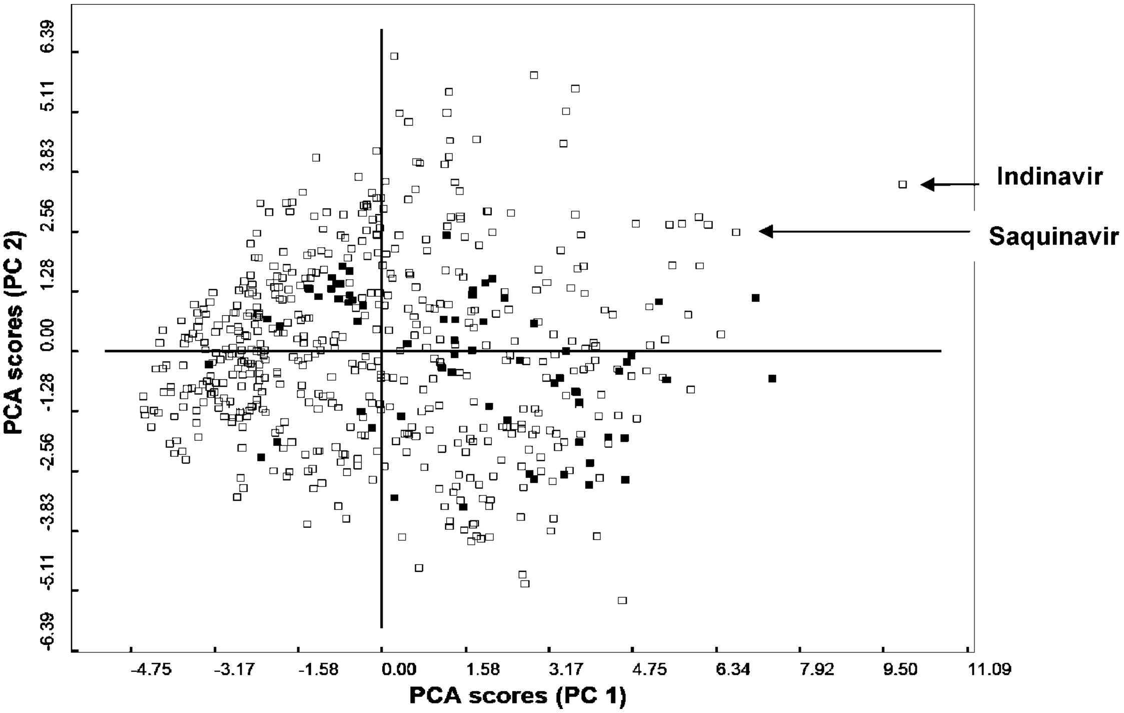


Figure 3. 2D plot of PCA showing the descriptor space of training set (open squares) and test set (filled squares) representing 561 compounds overall.

to identify the redundant structures in the data set, the compounds were clustered on the basis of the Tanimoto coefficients using bit fingerprints (166 MACCS structural keys) with a cutoff of 0.85. The 561 compounds can be grouped in 474 clusters where each cluster contains no more than four compounds with a high similarity, except one cluster which contains 14 sertindole analogues. With a cutoff of 0.60, the data set can be divided in 322 clusters whereas for a cutoff of 0.90, 496 clusters are generated. Thus, the data set is evenly distributed and there are no significant biased sample points.

Finally, we considered the chemical space defined by the GRIND descriptors. GRIND descriptors were calculated for the entire set of hERG blockers and nonblockers. The structural variation of the data set was analyzed with principal component analysis (PCA)[59] performed on the complete set of GRIND descriptors calculated for the compounds that comprised the training and test sets. The first two components explained around 40% of the structural variation of the data set. Figure 3 shows that only one compound could be defined as a structural outlier. This compound is indinavir, a FDA approved inhibitor of the human immunodeficiency virus (HIV) protease,[60] from which no hERG inhibition was detected. This drug contains two ionizable basic nitrogen atoms and is quite unique in this set. Only saquinavir, another HIV protease inhibitor which is considered a weak hERG blocker,[61] shows similar structural features. This compound is also found close to indinavir on the PCA map. Neverthe-

[59] Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.

[60] Weber, J.; Mesters, J. R.; Lepsik, M.; Prejdova, J.; Svec, M.; Sponarova, J.; Micochova, P.; Skalicka, K.; Strisovsky, K.; Uhlikova, T.; Soucek, M.; Machala, L.; Stankova, M.; Vondrasek, J.; Klimkait, T.; Kraeusslich, H. G.; Hilgenfeld, R; Konvalinka, J. Unusual binding mode of an HIV-1 protease inhibitor explains its potency against multidrug-resistant virus strains. *J. Mol. Biol.* **2002**, *324*, 739–754.

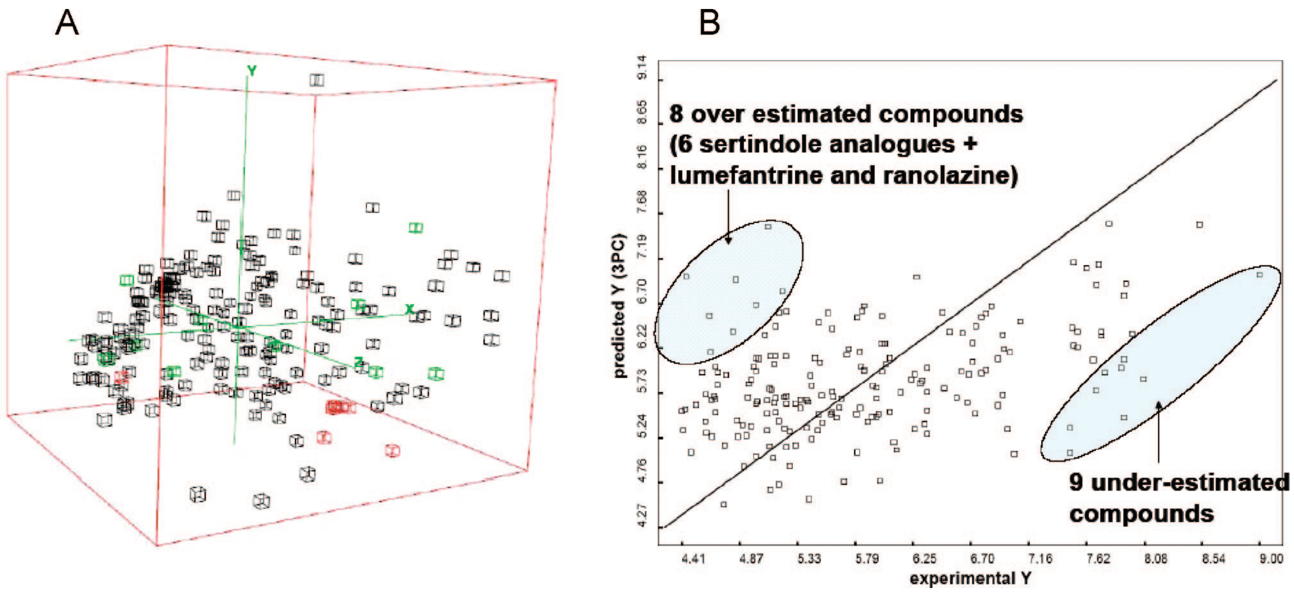


Figure 4. (A) 3D plot of PCA showing the distribution of the sertindole analogues (red squares) and underestimated compounds (green squares) within the 192 compounds. (B) Predicted versus experimental pIC_{50} for the 192 compounds. Both clusters in blue represent compounds removed to optimize the PLS model.

less, the training set and test set share similar chemical spaces and accordingly, we decided to use the entire set of compounds for our classification.

Analysis of the Blockers with GRIND Descriptors. Initially, we attempted to model the data set using only compounds with a quantitative activity (IC_{50} or K_i) less than 40 μM. The idea behind this assumption was to see if common pharmacophoric descriptors can be retrieved for the hERG inhibitors. With a total set of 192 compounds for which the pIC_{50} was defined, the PLS analysis resulted in an $r^2 = 0.34$ with three latent variables (LVs). The cross-validation of the model using the leave-one-out (LOO) method yielded a q^2_{LOO} value of 0.07 with a standard deviation error for the activity prediction (SDEP) around 0.95. It seems that most of the sertindole analogues are overestimated with a SDEP up to 1.5. Looking at the PCA (Figure 4A), most of the sertindole analogues (red squares) are located in the same area. The only one quite far from the others is a sertindole analogue with a quite large modification (absence of ethyl cyclic urea substituent). Probably the introduction of a bias value for the sertindole analogues would optimize the model. Other compounds, like gbr12909, terfenadine, E_4031, vinpocetine, droperidol, H345-32, cisapride, or clemastine are in opposite underestimated with the PLS model (Figure 4B). For these compounds, the underestimation is not directly related to the structural specificity like with sertindole analogues. On the PCA plot (Figure 4A), these molecules (green squares) are spread all over the map. Instead, it seems the high biological activity defined for these compounds reduces the performance of the model. For example, an IC_{50} of 8.4 nm was associated with terfenadine in this study. Another publication has referred instead to an IC_{50} of 330 nm,[57] which is around 40 times loss of activity. The same problem can be seen with the vinpocetine molecule. An IC_{50} of 32 nm was considered in our study, whereas recently an IC_{50} of 130 nm was reported.[62] Clearly, the introduction of uncertainty in the biological activity reduces the performance of the model, and we expected to get a poor predictive model with weak correlation between GRIND descriptors and the activity. However, removing these sertindole analogues and these underestimated compounds, we obtained a significantly better model ($r^2 = 0.57$, $q^2_{LOO} = 0.41$, and SDEP $= 0.72$ with three latent variables). So, the model based on 175 compounds showed some features that correlated with the inhibition activity. The importance of these pharmacophoric descriptors is depicted in Figure 5. The descriptor with the highest PC coefficient is the GRIND descriptor 22-18. It is related to the presence of two hydrogen bond donor atoms placed 9 Å apart and results from the charged basic nitrogen atom which has preferential interactions with this type of probe. The descriptors related to the presence of a hydrogen bond donor and a hydrogen bond acceptor placed 9.5 or 6 Å apart (descriptor 23–19 and 23–12, respectively), but also the descriptors related to the presence of the hydrophobic area and hydrogen bond donor group at 9 and 16 Å apart (descriptors 12–18 and 12–32, respectively), are significantly represented in the set of blockers. Finally, a distance of 10 or 17.5 Å between a hydrogen bond donor and one of the edges of the same molecule is also highly correlated to the

(61) Anson, B. D.; Weaver, J. G.; Ackerman, M. J.; Akinsete, O.; Henry, K.; January, C. T.; Badley, A. D. Blockage of HERG channels by HIV protease inhibitors. *Lancet* **2005**, *365*, 682–686.

(62) Yunomae, K.; Ichisaki, S.; Matsuo, J.; Nagayama, S.; Kukuzaki, K.; Nagata, R.; Kito, G. Effects of phosphodiesterase (PDE) inhibitors on human ether-a-go-go related gene (hERG) channel activity. *J. Appl. Toxicol.* **2007**, *27* (1), 78–85.

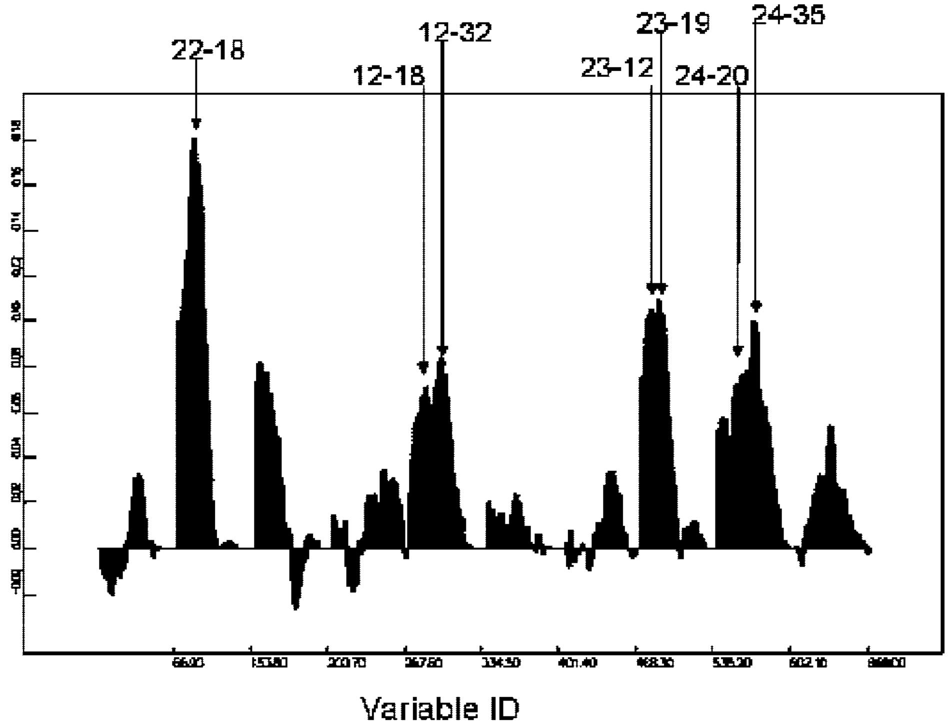


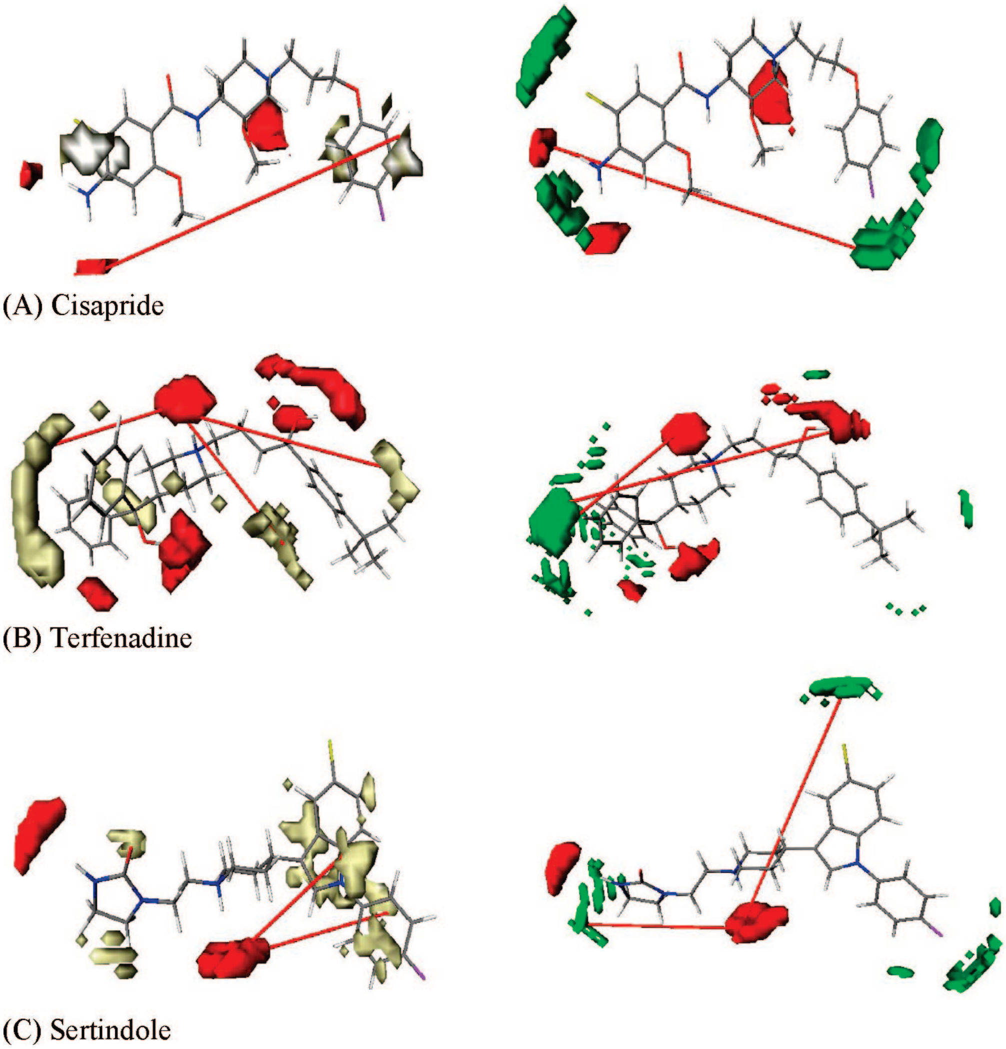
Figure 5. PLS coefficients profile of the pharmacophoric descriptors. The PLS plot coefficient highlights the GRIND variables which are directly (positive values) or inversely (negative values) correlated to the biological activity.

variation of the experimental data (descriptor 24–20 and 24–35, respectively).

Comparing our pharmacophore model with the previous model obtained by Cianchetta et al.,[25] we arrived at the same conclusion regarding the presence of a hydrophobic feature, the importance of a hydrogen bond donor, and also a hydrogen bond acceptor to bind to the hERG channel. Although the set of compounds and the structural conformation used in both studies is different, the optimal distance between the MIFs produced by a hydrophobic and a H-bond donor group was on the same order (16 Å in our study, while the distance was between 14 and 21Å depending on the presence of any basic nitrogen atom in the previous pharmacophore models). On the other hand, the optimal distance between the hydrogen bond donor and the edge of a molecule (descriptors 24) is much longer in the previous model (between 23 and 26 Å). The use of the hERG cavity to design the structural conformation of the molecules in our study probably gives a more bulky conformation compared the geometries generated in the other models.

The advantage of using GRIND descriptors is that these can be visualized for each molecule. In Figure 6, the pharmacophoric representation for cisapride, terfenadine, and sertindole, well-known hERG blockers, confirms the relevant features of the charged basic nitrogen and hydrophobic groups for binding to the hERG channel. It is likely that the generation of internally consistent hERG data sets would improve the performance of the model. Nevertheless, the GRIND descriptors seem to be able to capture the important features of the hERG blockers, and thus, we decided to implement them in association with a SVM method in a classification model which could distinguish between hERG blockers and nonblockers.

Classification with a Support Vector Machine. The entire data set comprising 495 drugs was separated into hERG blockers and nonblockers using 40 μM as threshold value. For some blockers defined by Aronov,[16] we did not find a corresponding inhibitor activity. Accordingly, only 476 compounds were considered at threshold values of 1, 5, 10, 20 and 30 μM, respectively. As all the models present a very high overall accuracy (up to 94% which represents 26 compounds misclassified), we focused our analysis on the LOO cross-validation, which is more sensitive to the different setup. The results of the classification models obtained from SVM are presented in Figure 7. In general, the quality of the models is not too sensitive to the threshold value for the linear models and seems to overfit the nonblockers set of the model. Instead, the nonlinear SVM method seems to be more general and gives slightly better results for a threshold of 20, 30, and 40 μM, respectively. Overall, the threshold value at 1 μM for a linear SVM yields the best result with 86% of correct prediction, but the performance of the model is overestimated by the specificity of the model with a MCC around 0.21. This model is clearly a good classifier of hERG nonblocker (392/422 correctly predicted in LOO) with a F-measure up to 0.90, but the sensitivity of the model (prediction of hERG blockers among hERG blockers) is rather low (18/54) with a F-measure around 0.30. The sensitivity is even worse for the nonlinear SVM with only 9/54 hERG blockers correctly predicted. The reason for the overestimation of hERG nonblockers is that the data are hardly separable with the support vectors from both classes.



(A) Cisapride

(B) Terfenadine

(C) Sertindole

Figure 6. Pharmacophoric element (red for O carbonyl probe, white for dry probe (hydrophobic), and green for TIP probe) for cisapride (A), terfenadine (B), and sertindole (C) identified by the ALMOND model.

In this extreme case, more hERG nonblockers are captured by support vectors and the margin between the two classes must be shifted to the hERG nonblockers. Consequently, many blockers will be misclassified as nonblockers.

Should the goal be the same as in drug discovery (i.e., to seek actives), then this appears to be useful. However, since *safety* is the primary concern and all drugs that cause QT prolongation and TdP need to be identified, much more data appears to be needed on many more drugs in order to train the model. Such models can be improved to greater accuracy at the cost of increasing the number of false positives.

For a threshold of 40 μM, the overall accuracy is close to 74% with a linear SVM (MCC of 0.40). The specificity of the linear model is good with 283 of 343 nonblockers correctly classified (*F*-measure of 0.81), and the sensitivity also becomes better with 83 of 152 blockers correct predictions (*F*-measure of 0.57). The best sensitivity is seen for the nonlinear SVM models. For a threshold of 30 μM, 94 of 129 blockers are correctly predicted. The sensitivity is also good at a threshold of 20 and 40 μM with more than 70% correctly classified. The specificity is also in the same order for these three thresholds, which means that these models seem to be more general.

In our data set, 359 of 561 compounds have a basic nitrogen atom, and of these, 197 are considered hERG blockers. It is well-known that a basic nitrogen atom is characteristic of most known hERG blockers. However, the effect of the basic center is modulated by other structural features like the surrounding of bulky hydrophobic groups. Of the 24 compounds misclassified in the linear SVM model at a threshold of 40 μM, 17 hERG blockers were predicted as nonblockers, and 11 of these have a tertiary amines or piperazine, like for example, tolterodine or vesnarinone. Sparfloxacin is also predicted as a nonblocker. This antimicrobial is the least potent hERG blocker in the quinolone antibacterial class in our training set with an IC_{50} value at 18 μm. This compound can be considered as a weak inhibitor and on the borderline of this model. Other antibacterials like ciprofloxacin, moxifloxacin, norfloxacin, and ofloxacin are considered not to cause TdP, though with some risk to congenital QT syndrome. The same remark can be applied for cocaine, defined as a blocker in the training set and which

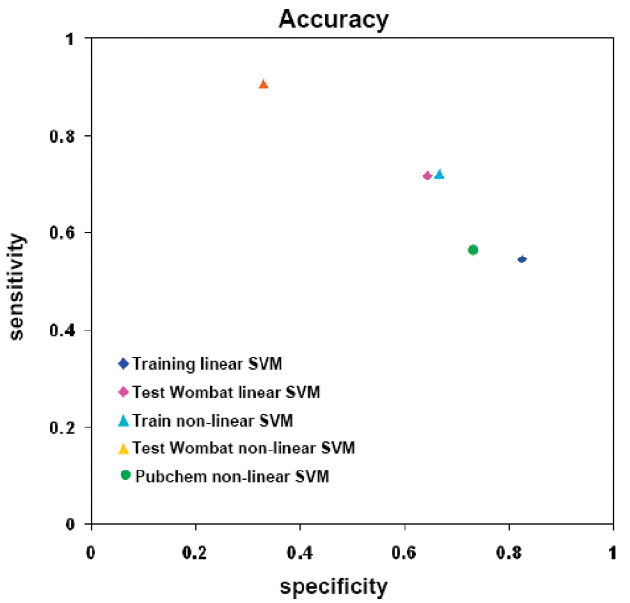


Figure 7. Sensitivity and the specificity are represented for the training set and test set for a 40 μM threshold. The blue diamonds correspond to the accuracy of the training in leave-one out, with the linear SVM. The cyan triangles represent the nonlinear SVM. The pink diamond and the orange triangle are, respectively, the accuracy on the WOMBAT-PK data set. The green circle represents the PubChem accuracy with the nonlinear SVM method.

is predicted as nonblocker. This anesthetic drug like atropine and cocaethylene is annotated as blocker in our training set, whereas ecgonine and its derivatives which also contain a tropane nucleus (like cocaine) are considered nonblockers. More information is needed for these classes of compounds to improve the model.

The models were tested using an external set of 66 compounds with hERG information extracted from the WOMBAT-PK database. The nonlinear model with threshold at 40 μM shows the best overall accuracy (72% of the compounds are correctly classified). The prediction of the blocker is rather good (85% (47/55) of the blockers correctly classified and F-measure of 0.86). On the contrary, prediction for nonblockers is only 36% (4 of 11 compounds correctly classified and F-measure of 0.34). In this external set, five of the wrong predictions are weak blockers (between 20 and 40 μM) and come from the same source.[63] All of them have an N-oxide group that should affect the hERG blockage, but the model is based on specific pharmacophoric groups and the optimal distance between them. The use of a specific conformation may alter the prediction of the model when they are weak inhibitors. Thus, the 3D conformation is quite an important element to take into account and the conformations captured by the docking can alter the prediction of a model, especially when they are considered as weak inhibitors.

The small size of the nonblocker set in the test set is probably not optimal for evaluating the performance of the models and testing on a large independent set would be preferable. Recently a hERG bioassay data set has become available on the PubChem database. The PubChem data set is divided in 248 active and 1700 nonactive compounds for a total of 1948 compounds providing an expanded test set. Among the 248 actives, 57 compounds are defined as openers and the rest (191 compounds) as blockers. After removal of 14 redundant compounds which overlap with the training set (4 blockers and 10 nonblockers), and using the same protocol as we described previously, we tested the nonlinear SVM model at a threshold of 40 μM, as it showed the best performance in general. Among the 57 openers, 55 were predicted as nonblockers. It was argued recently that the activation of hERG channels can be achieved by mechanisms of action which can be different from the hERG blockers.[64] As our model is defined to discriminate between blockers and nonblockers, it is suitable to remove these compounds from the blockers set. For the rest of the active set, 107 of 187 compounds were correctly predicted, with a F-measure around 0.30. For the nonblockers, 1271 of 1690 compounds are correctly predicted with a F-measure of 0.83. SVMs do not have a distance-to-model estimate and are unlikely to cope well with "unclassifiable" input objects. In other words, the SVM model may be less sensitive to uncertainty, but it is also likely to result in gross errors for unusual compounds.[58] To check the reliability of the method in our study, we use the model for the prediction of arsenic trioxide, a well-known QT prolongator (40% patients with 500 ms QTc), which seems not to be directly related to the hERG blockage (>300 μm).[65] The model predicts arsenic trioxide as a non blocker, which is in agreement with the recent publication.

Finally, we compared the performance of the SVM classification method on hERG to a random forest (RF) decision tree.[66,67] The leave-one-out correlation coefficient was in the same order of magnitude as the SVM model with 73% (361/495) correct prediction, but the performance of this model on the external PubChem data set yields a lower prediction for the blockers, with only 70 of 187 compounds correctly predicted. We also included the VolSurf hERG

(63) Friesen, R. W.; Ducharme, Y.; Ball, R. G.; Blouin, M.; Boulet, L.; Coté, B.; Frenette, R.; Girard, M.; Guay, D.; Huang, Z.; Jones, T. R.; Laliberté, F.; Lynch, J. J.; Mancini, J.; Martins, E.; Masson, P.; Muise, E.; Pon, D. J.; Siegl, P. K. S.; Styhler, A.; Tsou, N. N.; Turner, M. J.; Young, R. N.; Girard, Y. Optimization of a tertiary Alcohol Series of phosphodiesterase-4 (PDE4) Inhibitors: Structure-Activity Relationship related to PDE4 inhibition and Human Ether-a-go-go related gene potassium channel binding affinity. *J. Med. Chem.* **2003**, *46*, 2413–2426.

(64) Casis, O.; Olesen, S. P.; Sanguinetti, C. Mechanism of Action of a novel Human ether-a-go-go-related gene channel activator. *Mol. Pharmacol.* **2006**, *69*, 658–685.

(65) Katchman, A.; Koerner, J.; Tosaka, T.; Woosley, L.; Ebert, S. N. Comparative evaluation of HERG currents and QT intervals following challenge with suspected torsadogenic and nontorsadogenic drugs. *JPET* **2006**, *316*, 1098–1106.

(66) Breiman, L. Bagging Predictors. *Machine Learning* **1986**, *24*, 123–140.

(67) Aha, D. Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. *Int. J. Man-Mach. Studies* **1992**, *36* (2), 267–287.

classifier[3] in our comparative study. Only 46% of our training model was correctly predicted and about 69% for the PubChem data set (*F*-measure of 0.22 and 0.80 for the blockers and nonblocker, respectively). We also observed that in both methods the prediction of the hERG blockers was rather low, with only 37% and 47% blockers correctly predicted. Thus, it seems that our model has a better performance on the prediction of blockers compared to other methods, which can be used in the filtering of potential hERG channel inhibitors.

Conclusion

In conclusion, we have developed a method that can be used for the rapid evaluation of cardiac toxicity liability via hERG inhibition in the context of virtual screening of compounds libraries. The model is based on 3D conformations−which are likely to be available in both ligand-based and structure-based virtual screening, and combines pharmacophoric elements to capture important features and optimal distances with a support vector machine classifier. The best model was characterized by a threshold of 40 μM for a large and diverse set of compounds representing different classes of drugs and can be useful for filtering compound libraries for hERG blockers and flag them at an early stage of the drug discovery process.

Abbreviations Used

CHO, Chinese hamster ovary cells; COS, *Cercopithecus aethiops* cells; FDA, Food and Drug Administration; GRID, 3D descriptors; GRIND, pharmacophoric descriptors; HEK, human embryonic kidney cells; hERG, human ether a go-go related gene; IC_{50}, inhibitory concentration of 50%; K_i, inhibition constant; LOO, leave-one-out; LV, latent variable; MACCs, structural key fingerprint; MIF, molecular interaction fields; PCA, principal component analysis; PC, principal component; pIC_{50}, $-\log IC_{50}$; PLS, partial least squares; Q^2cross-validation, correlation coefficient; QSAR, quantitative structure–activity relationship; R^2, correlation coefficient; RBF, radial bias function; RF, random forest; SDEP, standard deviation error for the activity prediction; SVM, support vector machine; TdP, Torsades de Pointes; XO, *Xenopus laevis* oocytes.

Acknowledgment. This study was supported by the Program Commission on Nanoscience, Biotechnology and IT (NABITT) and the Danish Research Council for Technology and Production Sciences grant, by the NIH Molecular Libraries Initiative award U54 MH074425-01 (T.I.O.). We are grateful to Svava Ósk Jónsdóttir for her help and discussions, Gabriele Cruciani for use of the GRID, VolSurf, and Almond programs, and Chris Workman for careful reading of the manuscript.

Supporting Information Available: A list of the 561 compounds used in this study is presented with structures in SMILES format and IC_{50} values of the known compounds and the binary classification obtained for a 40 μM threshold. A table of the results for the different thresholds is also given. This material is available free of charge via the Internet at http://pubs.acs.org.

MP700124E