Pergamon

SCIENCE DIRECT

Bioorganic & Medicinal Chemistry Letters 13 (2003) 2773–2775

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

# Prediction of hERG Potassium Channel Affinity by Traditional and Hologram QSAR Methods

György M. Keserü*

*Computer Assisted Drug Discovery, Gedeon Richter Ltd., PO Box 27, H-1475 Budapest, Hungary*

Received 19 February 2003; revised 10 April 2003; accepted 9 May 2003

**Abstract**—Traditional and hologram QSAR (HQSAR) models were developed for the prediction of hERG potassium channel affinities. The models were validated on three different test sets including compounds with published patch-clamp $IC_{50}$ data and two subsets from the World Drug Index (compounds indicated to have ECG modifying adverse effect and drugs marked to be approved, respectively). Discriminant analysis performed on the full set of hERG data resulted in a traditional QSAR model that classified 83% of actives and 87% of inactives correctly. Analysis of our HQSAR model revealed it to be predictive in both $IC_{50}$ and discrimination studies.


Drug-induced QT interval prolongation, that might result in sudden cardiac death, has been identified as a critical side effect for numerous drugs.[1] Compounds with this cardiotoxic potential should therefore be identified in an early stage of the screening cascade that requires a reliable, preferentially in vitro assay. $I_{Kr}$ potassium channels encoded by the human ether-á-go-go related gene (hERG) are responsible for the normal action repolarization in the heart and serve as a therapeutic target for class III antiarythmics.[2] Inhibitors of hERG $K^+$ channels, however, can cause long QT syndrome that suggests the use of this system as an in vitro cardiotoxicity assay.[3] In addition to the most reliable whole cell patch-clamp investigations hERG $K^+$ channel affinity can be measured by the Rb-efflux based functional test, as well as by the $^3$H-dofetilide binding assay.[4,5] Considering the multiple binding sites of hERG $K^+$ channels[6] functional assays are preferred for screening. Due to the limited capacity of functional assays, particularly the low throughput of the whole cell patch-clamp electrophysiology, we are looking for a fast in silico screen for the prediction of hERG affinity of large compound libraries.

Recent approaches of virtual screens applied for hERG $K^+$ channels involve pharmacophore mapping[7] and CoMFA analysis.[8] Both approaches, however, assumed that compounds bind to the same binding site of the channel using similar binding modes.

Most hERG channel blockers that have been studied in detail (e.g. dofetilide, cisapride that suggested to interact with the inactivated state) are high-affinity ligands and exhibit limited voltage-dependent block. In contrast, low-affinity ligands (chloroquine, quinidine suggested to interact with the activated state) are characterized by significant voltage-dependent kinetics and steady-state effects.[9] Different characteristics of inhibitors suggest that the binding affinity of a drug might vary as a function of the channel state (activated/inactivated)[10] and prevent the application of a single pharmacophore model. Furthermore, one of the known binding sites of the inactivated hERG channel is located in the extended inner vestibule[11] and therefore the binding modes of high-affinity drugs could be different. Uncertainties of pharmacophore-based approaches prompted us to develop new, more general QSAR models.

Inhibitors of hERG $K^+$ channel were extracted from Fenichel's database[12] and ref 13. Compounds with $IC_{50}$ values for inhibition of hERG $K^+$ channels expressed in mammalian cells (HEK, CHO, COS, neuroblastoma cells) were selected. All $IC_{50}$ data were measured at 22 °C. Descriptors were calculated by CPSA and Biobyte programs in Sybyl 6.9[14] (29 descriptors) and Volsurf 3.04[15] (72 descriptors). The following strategy was

*Corresponding author. Tel.: +36-1-431-4605; fax: +36-1-432-6002; e-mail: gy.keseru@richter.hu

0960-894X/03/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0960-894X(03)00492-X

**Table 1.** Measured and predicted hERG $pIC_{50}$ data of training and test (bold) sets

| Compd | Measured | Traditional Qsar | Hologram Qsar | Compd | Measured | Traditional Qsar | Hologram Qsar |
|---|---|---|---|---|---|---|---|
| | Exp. $pIC_{50}$ | Pred. $pIC_{50}$ | Pred. $pIC_{50}$ | | $pIC_{50}$ | Pred. $pIC_{50}$ | Pred. $pIC_{50}$ |
| Alosetron | 5.49 | 4.81 | 5.48 | Mibefradil | 5.84 | 7.18 | 5.89 |
| Amiodarone | 5.00 | 5.25 | 4.89 | Mizolastine | 6.36 | 6.15 | 6.52 |
| Amitriptyline | 5.00 | 5.52 | 5.51 | Moxifloxacin | 3.89 | 3.57 | 3.97 |
| Astemizole | 8.00 | 7.59 | 8.00 | Nicotine | 3.61 | 3.75 | 4.14 |
| Azimilide | 5.85 | 6.17 | 6.22 | Nifedipine | 4.30 | 3.78 | 4.87 |
| Bepridil | 6.26 | 5.46 | 5.91 | Nitrendipine | 5.00 | 4.61 | 4.83 |
| Carvedilol | 4.98 | 6.94 | 4.63 | Norclozapine | 5.35 | 5.75 | 5.80 |
| Cetirizine | 4.52 | 5.63 | 4.39 | Ofloxacin | 2.85 | 3.76 | 3.02 |
| Chlorpheniramine | 4.68 | 5.66 | 4.68 | Ondansetron | 6.09 | 5.04 | 5.74 |
| Chlorpromazine | 5.83 | 5.92 | 6.13 | Perhexiline | 5.11 | 6.19 | 5.18 |
| Ciprofloxacin | 3.02 | 3.58 | 3.09 | Pimozide | 7.30 | 7.07 | 7.24 |
| Cisapride | 7.40 | 6.37 | 7.35 | Quinidine | 6.49 | 5.34 | 6.71 |
| Clarithromycin | 4.23 | 4.50 | 4.23 | Risperidone | 6.82 | 5.71 | 6.90 |
| Clozapine | 6.49 | 5.48 | 6.01 | Sertindole | 8.00 | 6.89 | 7.86 |
| Cocaine | 5.14 | 4.68 | 5.11 | Sildenafil | 5.48 | 5.73 | 5.43 |
| Ziprasidone | 6.92 | 6.44 | 6.99 | Sparfloxacin | 4.74 | 4.69 | 4.79 |
| Desipramine | 5.86 | 5.97 | 5.56 | Terfenadine | 6.70 | 6.98 | 6.77 |
| Diltiazem | 4.76 | 5.61 | 4.54 | Terikalant | 6.60 | 6.80 | 6.91 |
| Diphenhydramine | 4.57 | 4.49 | 4.35 | Thioridazine | 6.44 | 5.54 | 6.45 |
| Disopyramide | 4.04 | 4.14 | 3.95 | Verapamil | 6.85 | 6.88 | 6.47 |
| Dofetilide | 8.00 | 7.50 | 8.21 | Vesnarinone | 5.96 | 5.87 | 5.94 |
| Dolasetron | 4.92 | 6.16 | 4.79 | **A-56268** | 4.48 | 5.36 | 4.23 |
| E4031 | 7.70 | 6.34 | 7.65 | **Citalopram** | 5.40 | 5.70 | 4.67 |
| Epinastine | 4.00 | 5.47 | 4.02 | **Desmetheylastemizole** | 9.00 | 7.29 | 7.99 |
| Gatifloxacin | 3.89 | 4.07 | 3.83 | **Droperidol** | 7.49 | 6.17 | 7.04 |
| Grepafloxacin | 4.30 | 5.02 | 4.16 | **Flecainide** | 5.41 | 6.26 | 5.19 |
| Halofantrine | 6.70 | 7.73 | 6.66 | **Fluoxetine** | 5.82 | 5.54 | 5.33 |
| Haloperidol | 7.52 | 6.64 | 7.57 | **Hismanal** | 8.22 | 7.42 | 8.00 |
| Ibutilide | 8.00 | 7.26 | 7.26 | **MDL-74156** | 5.23 | 4.29 | 6.58 |
| Imipramine | 5.47 | 5.62 | 5.70 | **Mefloquine** | 5.25 | 6.26 | 5.78 |
| Ketoconazole | 5.72 | 6.28 | 5.82 | **Norastemizole** | 7.55 | 5.45 | 6.56 |
| Levofloxacin | 3.04 | 4.00 | 3.02 | **Olanzapine** | 6.74 | 5.24 | 6.19 |
| Loratadine | 6.77 | 6.71 | 6.74 | **RP-58866** | 6.70 | 5.84 | 6.91 |
| Mesoridazine | 6.49 | 5.64 | 6.61 | **Trimethoprin** | 3.62 | 3.75 | 4.39 |

persued to develop the new hERG affinity prediction method: (i) Retrieval of 68 drug-like structures with experimental, patch-clamp hERG $IC_{50}$ values from the literature (Table 1), (ii) division of 68 compounds into a 55 compound training set and a 13 compound test set while maintaining similar distributions of hERG activities in both sets, (iii) calculation of 99 potentially relevant descriptors, (iv) stepwise linear regression analysis with 99 descriptors and 55 training set compounds (using Statistica 6.0[16]), (v) testing of the derived equation on 13 test set compounds, (vi) testing on an additional 82 compound test set, namely molecules indicated to be 'ECG changer' as an adverse effect in World Drug Index[17] (WDI) under blind conditions. Stepwise linear regression analysis resulted in a QSAR model of acceptable quality ($R = 0.97$, $SSE = 0.82$, $F = 222.79$) with five descriptors including ClogP, molar refractivity (CMR), partial negative surface area (PNSA1), and two Volsurf descriptors W2 (polarizability) and D3 (hydrophobicity). Details of the QSAR model are shown in Table 2.

**Table 2.** Detailed description of the QSAR model

| Descriptor | Beta | SE beta | B | SE B |
|---|---|---|---|---|
| ClogP | −0.1099 | 0.0692 | −0.1557 | 0.0980 |
| CMR | −0.6736 | 0.0945 | −0.3571 | 0.0501 |
| PNSA1 | −0.2418 | 0.0861 | −0.0124 | 0.0044 |
| W2 | 0.0960 | 0.0406 | 0.5448 | 0.2306 |
| D3 | −0.0096 | 0.0358 | 0.0557 | 0.2064 |

Evaluation of the test set (Table 1) revealed this model to be moderately predictive ($R = 0.75$, $SSE = 0.98$, $F = 15.58$). Since we were not satisfied with the predictive power of this equation, descriptors of the model were further analyzed (Table 3).

Comparison of mean values of descriptors revealed that active and inactive subsets are different in molecular size, polarizability and hydrophobicity. Distributions of QSAR variables suggest a pausible interpretation of the model, that is liphophilic (ClogP, W2), highly polariz-

**Table 3.** Mean values of descriptors and their standard errors calculated in the hERG inactive and active datasets

| Descriptor | Inactive | | Active | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| ClogP | 2.539 | 0.380 | 4.493 | 0.287 |
| CMR | 10.195 | 0.450 | 16.226 | 4.667 |
| W2 | −0.126 | 0.166 | 0.163 | 0.165 |
| D3 | −0.383 | 0.154 | 0.495 | 0.143 |
| PNSA1 | 98.595 | 9.142 | 110.103 | 9.826 |

able (D3) compounds with basic character (PNSA1) and large in size (CMR) are likely to interact with the hERG channel. On the basis of these observations, we supposed that a predictive classification scheme can be developed. In fact, discriminant analysis performed on the full set of hERG data (compounds with $IC_{50}$ value smaller than 1 μM were considered as actives) resulted in a model that classified 83% of actives and 87% of inactives correctly. Coefficients of discriminant functions are collected into Table 4.

Prediction on ECG changer set of WDI demonstrated that this discriminant model recognizes potentially hERG-active compounds with high accuracy. Compounds of this dataset were classified into active or inactive groups depending on the highest classification score calculated by the corresponding functions. 67 out of the 82 (82%) ECG changer compounds were classified to be hERG active.

The limitation of our QSAR approach, however, is the slow calculation of the two Volsurf descriptors (W2 and D3) associated with the time-consuming generation of 3D structures for all the investigated compounds. Since high throughput in silico screens requires much faster prediction, we applied the hologram QSAR technique for the same set of data.

Hologram QSAR (HQSAR)[18] utilizes a special type of fragment fingerprints (holograms) as predictive variables of biological activity. In HQSAR, each molecule is divided into structural fragments that are counted in the bins of a fixed length array to form a molecular hologram. The bins are occupied by structural descriptors (independent variables) encoding compositional and topological molecular information. Then, QSAR model is generated through PLS regression by deriving a linear regression equation that correlates variation in structural information with variation in property data. Therefore this methodology avoids descriptor calculation, selection of conformers and the structural alignment that makes the use of 3D QSAR non-trivial. The method uses only 2D structure information and there is no necessity for molecular alignments. HQSAR models can thus be obtained more rapidly than models by other techniques and enables HQSAR readily applicable for both small and large data sets.[19]

Several HQSAR models with different hologram length were generated using the training set. The best HQSAR model was obtained with holograms containing 307 bins using six components in PLS [$R = 0.98$, $R(cv) = 0.80$, $SSE = 0.26$]. This model was first validated on the test set of 13 compounds (Table 1) and revealed to be highly predictive ($R = 0.90$, $SSE = 0.67$, $F = 55.74$). Next we applied the test set of 'ECG changer' compounds from WDI and identified 81% of these compounds to be hERG active when using hERG $IC_{50}$ of 1 μM as a borderline activity. This result is virtually identical to that obtained by the discriminant model, but the consumed processor time was reduced by a factor of 10. In our final test, we evaluated drugs indicated to be 'APPROVED' in WDI. Our HQSAR model predicted only 743 out of 4043 (18%) drugs to be hERG active indicating that the number of false positives is reasonable. These final tests revealed that the HQSAR model predicts hERG affinity with reasonable accuracy, 81% of potentially actives, but only 18% of potentially inactives was predicted to be active.

Results obtained by different test sets suggest that our HQSAR model might be a powerful in silico screen for drug discovery programs.

**Table 4.** Coefficients of discriminant functions obtained for the full hERG dataset

| Descriptor | Inactive | Active |
|---|---|---|
| ClogP | −5.7613 | −4.8274 |
| CMR | 10.1142 | 9.2627 |
| W2 | −35.3463 | −30.9344 |
| D3 | −5.9010 | −4.6170 |
| PNSA1 | 0.4130 | 0.3726 |
| Constant | −68.5500 | −59.5657 |

## References and Notes


1. Viskin, S. *Lancet* **1999**, *354*, 1625.
2. Crumb, W. J.; Cavero, I. *Pharm. Sci. Technol. Today* **1999**, *2*, 270.
3. Vandenberg, J. I.; Walker, B. D.; Campbell, T. J. *Trends Pharmacol. Sci.* **2001**, *22*, 240.
4. Sanguinetti, M. C.; Jiang, C.; Curran, M. E.; Keating, M. T. *Cell* **1995**, *81*, 299.
5. Biofocus Plc.
6. Sanchez-Chapula, J. A.; Navarro-Polanco, R. A.; Culberson, C.; Chen, J.; Sanguinetti, M. C. *J. Biol. Chem.* **2002**, *277*, 23587.
7. Cavalli, A.; Poluzzi, E.; De Ponti, F.; Recanatini, M. *J. Med. Chem.* **2002**, *45*, 3844.
8. Ekins, S.; Crumb, W. J.; Sarazan, D.; Wikel, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Ther.* **2002**, *301*, 427.
9. Sanchez-Capula, J. A.; Salinas-Stefanon, E.; Torres-Jacome, J.; Benavides-Haro, D. E.; Navarro-Polanco, R. A. *J. Pharmacol. Exp. Ther.* **2001**, *297*, 437.
10. Ficker, E.; Obejero-Paz, C. A.; Zhao, S.; Brown, A. M. *J. Biol. Chem.* **2002**, *277*, 4989.
11. Mitcheson, J. S.; Chen, J.; Lin, M.; Culberson, C.; Sanguinetti, M. C. *Proc. Natl. Acad. Sci. U.S.A* **2000**, *97*, 12329.
12. Fenichel, R. R. http://www.fenichel.net/QT%20stuff
13. Roche, O.; Trube, G.; Zuegge, J.; Pfimlin, P.; Alanine, A.; Schneider, G. *ChemBioChem.* **2002**, *3*, 455.
14. *SYBYL 6.9*; Tripos Inc.: 1699 South Hanley Road, St. Louis, MO, 63144, USA.
15. *Volsurf 3.0.7d*; Molecular Discovery Ltd.: 4 Chandos Street, London W1A 3AQ, UK.
16. *Statistica 6.0*; StatSoft, Inc.; 2300 East 14th Street, Tulsa, OK 74104, USA.
17. *World Drug Index*; Thomson Derwent: 14 Great Queen Street, London WC2B 5DF, UK.
18. Heritage, T. W.; Lowis, D. R. Molecular Hologram QSAR. In *Rational Drug Design*; Parrill, A. L., Reddy, M. R., Eds.; ACS Symposium Series 719; American Chemical Society: Washington, DC, 2000; p 212.
19. Tong, W.; Lowis, D. R.; Perkins, R.; Chen, Y.; Welsh, W. J.; Goddette, D. W.; Heritage, T. W.; Sheehan, D. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 669.