

Prediction of HERG affinity values using independent approaches

Elena Fioravanzo (a), Nicola Cazzolla (b), Lucia Durando (b), Cristina Ferrari (a), Massimo Mabilia (a), Rosella Ombrato (b).

(a) S-IN Soluzioni Informatiche Via Salvemini 9, 36100, Vicenza, Italy (www.s-in.it)

(b) A.C.R.A.F. S.p.A, P.le Stazione I-00040, S. Palomba, Rome, Italy

Introduction

Drug-induced QT interval prolongation, as measured on the human electrocardiogram, was once considered a trivial physiological finding. Now it is believed that drug-induced QT interval prolongation, that has been identified as a critical side effect for numerous drugs, might result in sudden cardiac death. As a consequence, a number of prescription medications associated with QT prolongation have been removed from the market by the European pharmaceutical regulatory authority. The focus of many in vitro studies to date is the membrane-bound inward (rapid activating delayed) rectifier potassium channel (I_{Kr}) also known as the product of the human ether-a-go-go gene (hERG). Drugs or their metabolites may block this channel, thereby prolonging the QT interval and in some cases leading to the potentially life-threatening ventricular arrhythmia that may degenerate into ventricular fibrillation and sudden death. Notably, blockade of hERG K^+ channel forms the basis of the therapeutic effect of class III antiarrhythmic drugs, but for all other drugs, it is an unwanted side effect that must be detected as early as possible during drug development [1].

Since at present various in vivo and in vitro models for QT prolongation and subsequent arrhythmia exist but they may not be entirely predictive for humans, the availability of in silico methods in the early phase of drug development would dramatically increase the screening rate and would also lower the costs compared to experimental assay methods. The possibility of a computational hERG model to be used as a filter in the discovery process would add an extra dimension to lead optimization. Both a quantitative and a qualitative model would theoretically enable virtual selection of candidates with the lowest potential to cause hERG inhibition.

Recent studies on hERG K^+ channels involve pharmacophore mapping and CoMFA study. Both approaches, however, are based on the assumption that different compounds bind to the same binding site of the channel using similar binding modes. On the contrary, it is reasonable to assume that the binding affinity of a given compound may vary as a function of the channel states (activated/inactivated), and that structurally diverse molecules may adopt different binding modes. Such considerations are not compatible with a single pharmacophore model nor with a common alignment criterion.

In this study different computational approaches are used to predict hERG K^+ channel affinities.

Methods

Data set. 70 compounds with experimental hERG IC_{50} values were retrieved from the literature [2] (Table 1).

EVA descriptor. The derivation of the EVA descriptor has previously been described elsewhere [3] and only a brief description of the technique will be given here. The descriptor is derived from IR- and Raman-range molecular vibrational frequencies usually calculated through the application of a normal coordinate analysis (NCA) to an energy minimized structure. For a compound with N atoms there are $3N - 6$ (or $3N - 5$ for a linear structure such as acetylene) normal modes of vibration. Thus, except in the special case where each structure has the same number of atoms, the number of frequencies will be different for each structure; that is, the property is in non-standard form. A technique has thus been developed in order to standardize the property such that each compound is characterized by an equivalent-length descriptor. The frequency set for a given structure is projected onto a linear bounded frequency scale (BFS) covering a range from 1 to 4000 cm^{-1} . A Gaussian kernel of fixed standard deviation (s) is then placed over each and every eigenvalue. The BFS is then sampled at fixed increments of $L\text{ cm}^{-1}$ and the value of the resulting EVA descriptor at each sample point is the sum of the amplitudes of the overlaid kernels at that point. This procedure is repeated for each dataset compound and then combined to provide a matrix with M rows (compounds) and $4,000/L$ columns (descriptor variables). Typically, a descriptor set has been derived using a s of 10 cm^{-1} and an L of 5 cm^{-1} giving 800 descriptor variables. For a standard QSAR dataset the number of variables is thus much larger than M and Partial least square to Latent Structure (PLS) is hence used to provide a robust regression analysis. Energy minimization and normal coordinate analysis were carried out by means of Spartan'02 [4] employing Merck Force Field. Calculation of EVA descriptor from vibrational frequencies was carried out using the proprietary program EVA-02 (S-IN).

DRAGON is a software package for the calculation of molecular descriptors developed by Milano Chemometrics and QSAR Research Group. It allows calculation of more than 1600 molecular descriptors for thousands of molecules [5].

PASS (*Prediction of Activity Spectra for Substances*) [6,7] predicts the probability for any given compound to be active (P_a) or inactive (P_i) for each one of over 1000 biological activities, including pharmacological effects, mechanisms of action, mutagenicity, carcinogenicity, teratogenicity, and embryotoxicity. P_a and P_i values vary from 0 to 1, and their sum may be different than 1. PASS predictions are based on the analysis of structure-activity relationships for a training set including a great number of non-congeneric compounds with different biological activities, using the descriptor Multilevel Neighborhoods of Atoms (MNA). PASS training set consists of over 46,000 biologically active compounds: 16,000 are already launched drugs and 30,000 drug-candidates under clinical or advanced preclinical testing.

QikProp [8,9] has been developed by Prof. Bill Jorgensen at Yale University to rapidly predict ADMET properties of drug candidates. QikProp results have been fitted to datasets of drug-like molecules, based on 2-D and 3-D descriptors reflecting Monte Carlo simulation studies as well as experiment. QikProp predictions are calculation-based, as opposed to fragment based. Fragment-based methods can be problematic when they do not recognize parts of a structure or encounter unfamiliar fragment interactions, whereas QikProp will calculate properties based on the whole molecule. The advantage of this approach is that QikProp can be applied to new and unknown scaffolds.

SIMCA-P+ [10] Projections to Latent Structures (PLS) modeling has been used to investigate likely correlations between EVA and experimental pIC_{50} values and, respectively, descriptors generated by DRAGON and experimental pIC_{50} values. The optimal number of components in each PLS model was determined by SIMCA-P+ default cross-validation procedure. Variables selection was carried out on the basis of VIP parameter and coefficient values. All PLS models here reported were generated considering just the experimental values found in ref. 2(b). Initial models were generated using all 62 compounds - strong outliers were detected and then excluded employing PCA on each X data matrix. The best models were

further validated considering half of the compounds as training set and the rest as external test set. Training and test sets were generated by means of Onion/D-Optimal Design with the software MODDE [10].

Table 1. Molecules included in the data set.

Primary ID	No.	pIC ₅₀ *	pIC ₅₀ **	pIC ₅₀ ***	Primary ID	No.	pIC ₅₀ *	pIC ₅₀ **	pIC ₅₀ ***
2-Hydroxymethyl olanzapine	48	-	4.9	-	grepafloxacin	61	4.1	-	4.3
9-OH risperidone	29	-	5.9	-	halofantrine	19	6.7	6.7	6.7
A 56268	58	-	-	4.5	haloperidol	8	7.6	7.6	7.5
Alosetron	35	-	5.5	5.5	ibutilide	4	-	-	8.0
amiodarone	47	-	-	5.0	imipramine	36	5.5	5.5	5.5
amitriptyline	45	-	-	5.0	ketoconazole	34	-	-	5.7
Astemizole	1	9.0	-	8.0	levofloxacin	69	-	-	3.0
Azimilide	26	6.3	-	5.9	loratadine	17	6.8	-	6.8
Bepidil	25	6.3	-	6.3	mefloquine	41	-	-	5.3
Carvediol	49	-	-	4.9	mesoridazine	22	-	6.5	6.5
Cetirizine	56	-	-	4.5	mibefradil	31	5.8	-	5.8
chlorpheniramine	52	-	-	4.7	mizolastine	24	6.5	-	6.4
chlorpromazine	32	5.8	5.8	5.8	moxifloxacin	65	3.9	-	3.9
ciprofloxacin	70	-	-	3.0	N-desmethylozapine	57	-	4.5	-
Cisapride	2	8.2	8.2	7.4	nicotine	68	-	3.6	3.6
Citalopram	39	-	-	5.4	nifedipine	59	-	-	4.3
Clozapine	18	6.7	6.5	6.5	nitrendipine	46	-	-	5.0
Cocaina	42	5.2	-	5.1	norastemizole	9	7.6	-	7.6
desipramine	30	-	5.9	5.9	norclozapine	40	-	-	5.4
Desmethyl olanzapine	50	-	4.9		olanzapine	20	-	6.6	6.7
Diltiazem	51	4.8	4.8	4.8	ondansetron	27	-	6.1	6.1
diphenhydramine	55	-	-	4.6	perhexiline	44	5.1	5.1	5.1
disopyramide	62	-	-	4.0	pimozide	7	7.7	7.3	7.3
Dofetilide	5	7.9	-	8.0	quinidine	23	-	-	6.5
Dolasetron	43	5.2	4.9	4.9	risperidone	16	6.8	6.8	6.8
domperidone	15	6.8	-	-	sertindole	6	7.9	7.8	8.0
Droperidol	10	7.5	-	7.5	sildenafil	63	4.0	5.5	5.5
E-4031	3	8.1	7.7	7.7	sparfloxacin	54	4.6	-	4.7
Epinastine	64	-	-	4.0	terfenadine	12	6.9	6.7	6.7
fexofenadine	53	4.7	-	-	terikalant	21	-	-	6.6
Flecainide	38	-	-	5.4	thioridazine	11	7.5	-	6.4
Fluoxetine	33	-	-	5.8	trimethoprin	67	-	-	3.6
gatifloxacin	66	3.9	-	3.9	verapamil	13	6.8	6.8	6.9
glibenclamide	60	4.1	-	-	vesnarinone	28	-	6.0	6.0
granisetron	37	5.4	5.4	-	ziprasidone	14	6.8	6.9	6.9

* Experimental data from [2a] ** Experimental data from [2c] *** Experimental data from [2b]

Results and discussion

EVA. Molecules A-56268 and perhexiline were detected as outliers by an initial PCA on the whole X matrix and were excluded from the data set. Results are reported in Table 2.

DRAGON descriptors. Molecules A-56268 and nicotine were detected as outliers by an initial PCA on the whole X matrix and were excluded from the data set.

Different combinations of DRAGON descriptors were tested: the entire matrix, 1399 descriptors, and, alternatively, every single block of descriptors. Best results in terms of predictivity of models were obtained with 2D Autocorrelation [11].

More recently, SIMCA-P+ Orthogonal Signal Correction (OSC) algorithm was used to remove from X data matrices (e.g. EVA, RDF) information that is orthogonal to Y: better models, both in terms of Q^2 and predictive power towards test sets, were thus obtained that will be reported elsewhere.

Table 2. Results of PLS models.

Descriptors	X	PCs	Obj.	R2	Q2	SDEC	SDEP
Initial models							
DRAGON ALL	418	3	60	0.817	0.678		
DRAGON 2Dautocorr.	48	2	60	0.657	0.527		
EVA	182	2	60	0.665	0.533		
Final models (R2 Q2 SDEC for training and SDEP for test set)							
DRAGON ALL	329	3	30	0.831	0.703	0.523	0.939
DRAGON 2Dautocorr.	34	2	30	0.727	0.572	0.663	0.827
EVA	86	2	30	0.745	0.615	0.669	0.985

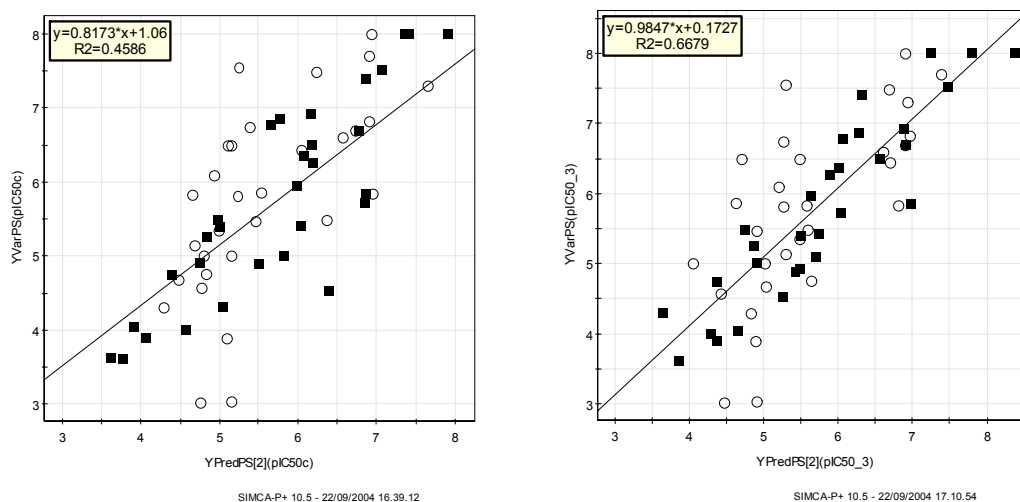


Figure 1. Scatter plot Yexp-Ypred. of final models, EVA on the left and DRAGON all on the right. Training set (box), test set (circle).

The program **PASS** was trained to predict the probability of hERG activity, using a set of molecules with pIC_{50} values greater than or equal to 5.0. Preliminary studies choosing a **PASS**

probability value (p_a) of 0.3 show that 75% of active molecules and, respectively, 73% of “inactive” molecules are predicted correctly, leading to 12 false negative and 6 false positive.

Results of **QikProp** predictions are reported in Figure 2. Considering 5 as the pIC_{50} threshold value between active and inactive compounds, 2 molecules are predicted as false negatives and 12 as false positives.

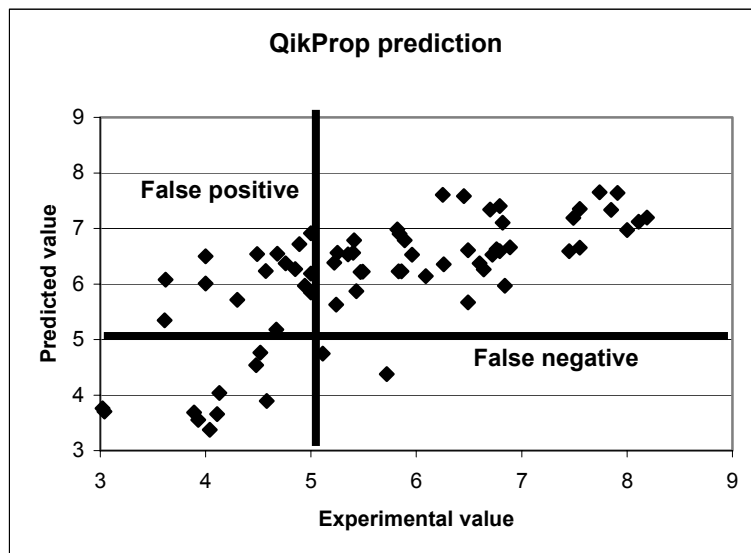


Figure 2. QikProp results.

References

- (1) (a) J.S. Mitcheson, J. Chen, M. Lin, C. Culberson, M.C. Sanguinetti *PNAS*, **2000**, 97, 12329. (b) D.A. Doyle, J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, *Science*, **1998**, 280, 69.
- (2) (a) A. Cavalli, E. Poluzzi, F. De Ponti, M. Recanatini, *J. Med. Chem.* **2002**, 45, 3844. (b) G.M. Keserü, *Bioorg. Med. Chem. Lett.* **2003**, 13, 2773. (c) S. Ekins, W.J. Crumb, R. Dustan Sarazan, J.H. Winkel, S.A. Wrighton, *J. Pharma. Exp. Therapeutics (JPET)* **2002**, 301, 427
- (3) (a) P. Jonathan, W.V. McCarthy, A.M.I. Roberts, *J. Chemometrics*, 10, **1996**, 189. (b) A.M. Ferguson, T. Heritage, P. Jonathan, S.E. Pack, L. Phillips, J. Rogan, P.J. Snaith, *J. Comput. -Aided Mol. Design*, 11 (**1997**) 143.
- (4) Spartan'02, Linux/Unix version, Wavefunction, Inc. www.wavefun.com.
- (5) DRAGON, Talete, srl, www.telemacus.it/talete.
- (6) V. Poroikov, D. Filimonov & Associates; www.ibmh.msk.su/PASS/.
- (7) (a) V. Poroikov, D. Filimonov (2001) *Computer-aided prediction of biological activity spectra. Application for finding and optimization of new leads. Rational Approaches to Drug Design*, Eds. H.-D. Holtje, W.Sipl, Prous Science, Barcelona, p.403-407 (b) S. Anzali, G. Barnickel, B. Cezanne, M. Krug, D. Filimonov, V. Poroikov, *J. Med. Chem.*, **2001**, 44, 2432-2437.
- (8) Schrödinger, L.L.C., www.schrodinger.com.
- (9) (a) E.M. Duffy, W.L. Jorgensen, *J. Am. Chem. Soc.* **2000**, 122, 2878-2888. (b) W.L. Jorgensen, E.M. Duffy, *Bioorg. & Med. Chem. Lett.* **2000**, 10, 1155-1158.
- (10) Umetrics AB, www.umetrics.com.
- (11) P. Broto, G. Moreau, C. Vandicke, *Eur. J. Med. Chem.*, **1984**, 19, 66-84.