

A simple approach discriminating cardio-safe drugs from toxic ones

Mizied Falah, Taher Nassar and Anwar Rayan*

Drug Discovery Informatics, QRC-Qasemi Research Center, Al-Qasemi Academic College, P.O.B. 124, Baka El-Garbiah 30100, Israel, Anwar Rayan - Email: a_rayan@qsm.ac.il; *Corresponding author

Received March 20, 2009; revised March 29, 2009; accepted April 05, 2009; published June 13, 2009

Abstract:

More than 130 FDA-approved drugs have been identified for now to prolong the QT interval and possibly lead to sudden cardiac death. Due to their toxic effect, some of these drugs have been withdrawn from the pharmaceutical market. In this study, we have formulated few rules to assess the ability to prolong QT interval and thereby discriminate between cardio-toxic and -safe drugs. These rules have clearly determined that cardio-toxic drugs are more likely to obey Lipinski rule of 5 and Oprea lead-like rule. Moreover, the cardio-toxic drugs have been found to have in common values of -0.5 to 6.5 log P, 1-5 nitrogen atoms, up to 4 oxygen atoms, 5-27 hydrophobic atoms, and 15-53 single bonds. Matthews Correlation Coefficient with the value of 0.6 was also attained and nearly 96% of the cardio-toxic drugs were successfully covered. Thus, despite the simplicity of this methodology, we have obtained interesting and informative results. The proposed set of these simple rules could be employed to infer cardio-toxicity or -safety for current and potential drugs. The present study will have important impact on decision making in the fields of drug development, molecule screening in biological assays, and other applications as well.

Keywords: Long QT Syndrome, cardio-toxic drug, prediction model

Abbreviations: LQTS = Long QT Syndrome; TDS = Torsades de Pointes; hERG = human ether-a-go-go-related gene; MCC = Matthews' Correlation Coefficient

Background:

Our thinking and perspective on drug development is now being revised due to serious cardiac side effects arising from certain drugs available in the market [1-2]. Endorsing or banning the process of developing a new drug is now more than ever subject to strict scrutiny. Meanwhile, drug-related patents and datasets of broader and higher quality are accumulating and becoming available in the database and as a result, computational approaches are developed in parallel aiming to predict cardiac toxicity of such drugs, their by-products, and the factors that induce them [3-5]. The focus of the current study is to verify certain drugs toxicity to the cardiac tissue and thereby differentiate non-toxic (safe) from toxic drugs. Such drugs lead to QT interval prolongation and causes fatal consequences.

Long QT syndrome (LQTS) is usually congenital although it can be acquired after administration of certain drugs known to affect the heart electrical system, prolong the QT-interval and cause a dangerous form of heart arrhythmia [6]. QT interval includes both the ventricular depolarization and repolarization intervals. Prolongation of QT interval induced by drugs is due to increased repolarization time through blockade of the outward potassium (K^+) current [7]. More than 80% QT-prolonging drugs have been shown to inhibit the K^+ channels encoded by the human ether-a-go-go-related gene (hERG) [1-2]. The hERG K^+ channel, a key target for QT-prolonging drugs, is a voltage-activated channel, expressed by multigene family forming tetrameric structure which consists of four identical α -subunits [8]. Each subunit consists of six α -helical transmembrane domains. Mutations occurred in the hERG localized in the cardiac

tissue causing loss of channel function leads to congenital LQTS [9].

Cardiac and non-cardiac drugs prolonging QT interval vary in their risk or cardiac toxicity [1-2, 10]. The most serious risk of such drugs is known by the French term Torsades de Pointes (TDP) which also called polymorphic ventricular tachycardia, a specific type of dangerous ventricular arrhythmia that could deteriorate to fatal ventricular fibrillation and results in sudden death. Although, some drugs carry a risk of TDP, others are only associated with that risk and lack substantial evidences for causing TDP. Other types of drugs carry the risk of TDP only with patients having congenital LQTS. Many non-cardiac drugs prolong the QT interval enhancing the risk of TDP as a side effect [1, 10]. TDP is usually self-limited arrhythmia causing heart shivering, dizziness, or even epilepsy and loss of consciousness [11]. TDP can also worsen and cause ventricular fibrillation eventually results in sudden death. Drug-induced TDP has been considered for the last decade to be the single most common cause of the withdrawal or restriction of the use of the corresponding marketed drugs [12]. The large number of non-cardiac drugs that prolong QT interval is rising, amongst are antihistamines, anti-psychotics, anti-bacterial, and others. Their potential side effects leading to cardio-toxicity and resulting in sudden death constitute a major concern in drug manufacturing and regulation. For that reason, the focus has been put onto current and potential drugs causing QT prolongation in the population [13]. As a result, several drugs have been withdrawn from the market or restricted in their use, whereas others due to their

QT prolongation and deadly cardiac consequences were not even granted the regulatory approval [12, 10, 14].

Considering the torsadogenic effect of current drugs and due to various measures taken by regulators as to drug withdrawal and restriction in the market, our view has been that while developing a drug, it is important to test its cardio-toxicity using computational tools that employ available information on drug structural characteristics. These tools could predict cardio-toxicity of a potential drug early in its development and thus based on available data could verify the LQTS risks saving significant time, labour and resources and more importantly avoiding loss of patients lives. In this research we have made efforts to define a set of counting rules to filter out compounds in databases that are not 'cardio-safe'. We have used characteristics of known cardio-toxic drugs in terms of molecular descriptors and optimized their ranges.

Methodology:

Learning sets

A corresponding search in the literature identified several drugs that exhibit QT-interval prolonging activity [15-18]. Rand Biotechnologies Ltd has conducted this search and provided us with categories of drugs as follows: 133 cardio-toxic drugs and 1500 potentially cardio-safe marketed drugs. In the current study we aim to develop a method that could provide simple rules to be used by medicinal chemists for verifying or designing cardio-safe molecules. For such purpose we have utilized 2D-descriptors which are based on simple whole-molecule properties e.g. molecular weight, H-bond donors, H-bond acceptors, log P, polar surface area, number of heavy atoms, number of hydrophobic atoms, number of rings, number of aromatic atoms and bonds, number of carbon, nitrogen, oxygen, sulphur and phosphor atoms.

Descriptors selection and processing

The decision as to which set of descriptors to be used for differentiating between cardio-toxic and -safe drugs is crucially important. We sought after the most significant set of descriptors from which guidelines for the design of safe compounds to be suggested. The selection has been performed as following: all descriptors were evaluated separately and the best discriminative descriptor was chosen to be the core. The second descriptor to be added to the core was selected from the rest descriptors while giving the best performance in discrimination. The process continued until we have five descriptors in the core.

We aim to construct a filter consisting of ranges of 5 descriptors that can differentiate well between cardio-toxic drugs and cardio-safe drugs. For this purpose, ranges of descriptors were optimized simultaneously in exhaustive search, by maximizing a function (Matthews' Correlation Coefficient, MCC) that considers each of the four possible outcomes for any drug – Positive, Negative, False Positive and False Negative (equation 1, **Supplementary material**). Higher MCC means better distinction. The division process of databases into training set and test set has been repeated 5 times with 80% of the cardio-toxic/cardio-safe drugs, while the remaining (20% of the databases) was used as a test set. Each time the division was performed by a random choice. The need for a combinatorial optimization of descriptors' ranges dictates the requirement to transform descriptor values into discrete ones. Some descriptors already have a discrete character, i.e., the numbers of Nitrogen atoms, H-bond acceptors etc, while others, such as molecular weight, polar surface area, etc., are continuous. The transformation to discrete character was limited to give 50 values for upper and lower limit ranges each.

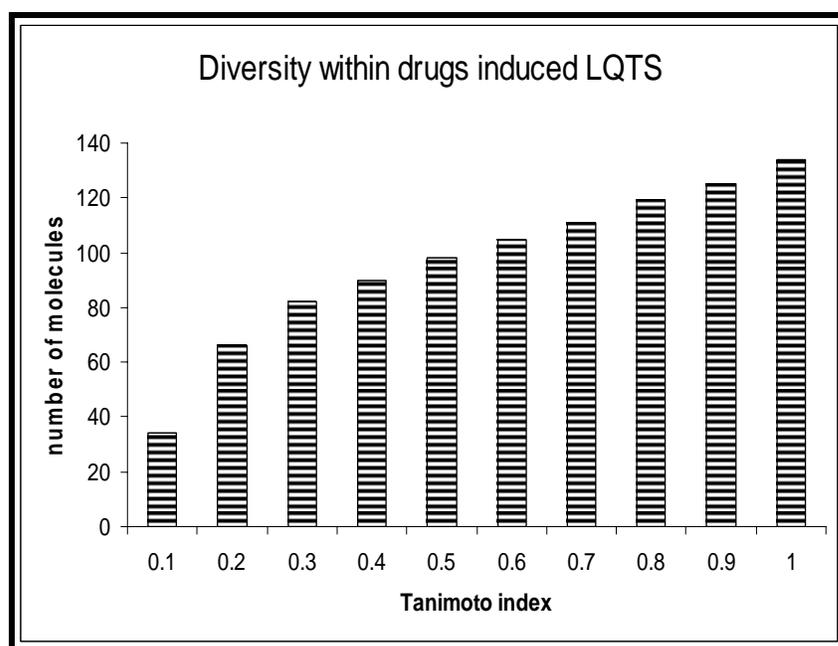


Figure 1: diversity within drugs prolonging QT interval.

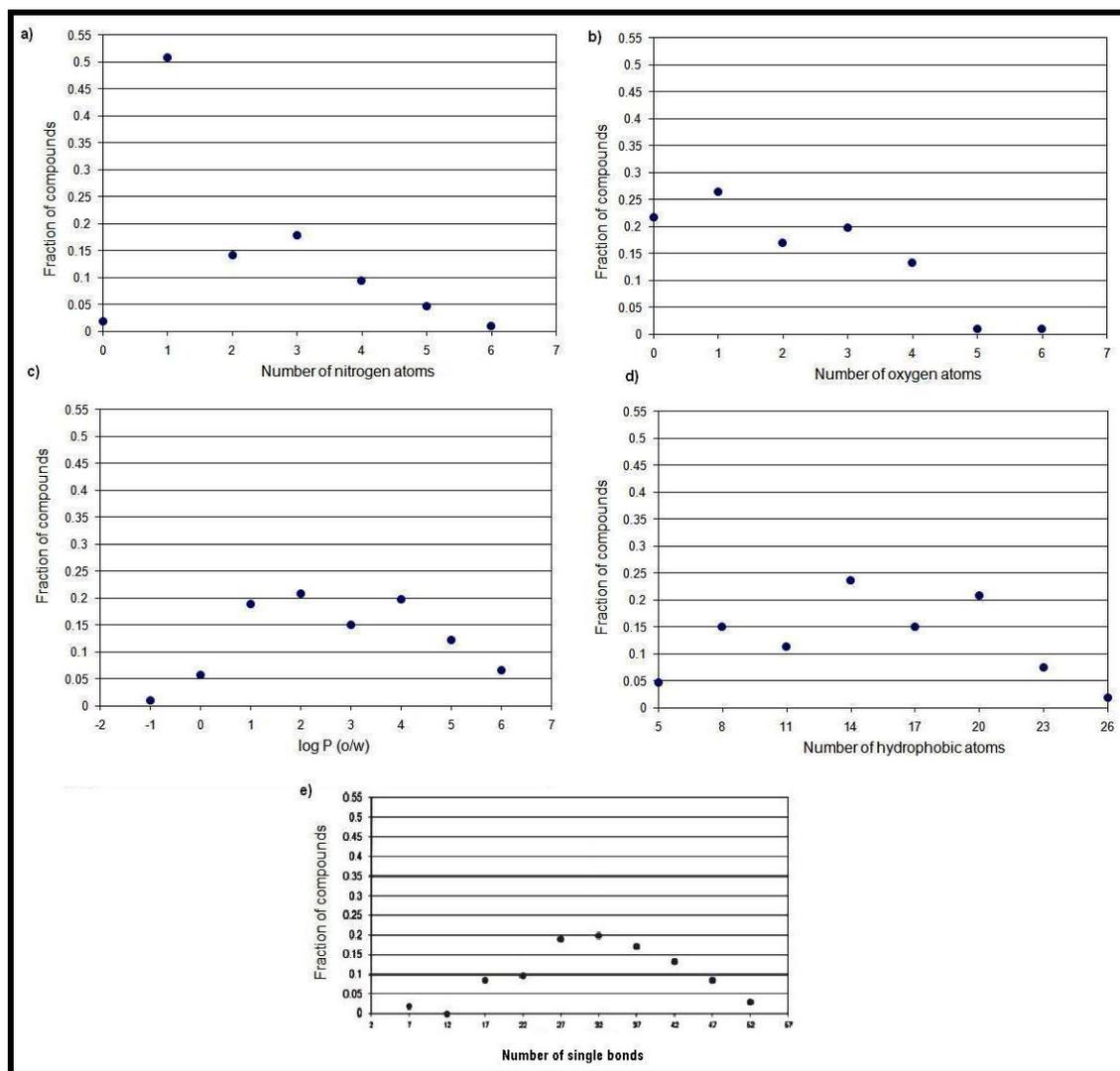


Figure 2a-e: Plots of the distribution patterns of number of nitrogen atoms, oxygen atoms, single bonds, hydrophobic atoms and log P in the LQTS drugs.

Prediction modelling

A set of rules is constructed by picking lower limit and upper limit for each descriptor. Each set has two values for each descriptor, constituting the range which is considered to be the “correct” one (by that set of descriptors) for cardio-safety. The “correctness” of this set is measured by its MCC value, described below. The constructed set of rules is applied to the drugs in the training sets to calculate the value of the scoring function, its Matthews Correlation Coefficient (MCC) (equation 1). An exhaustive search is performed for all combinations (around 250 million options) and the resulting sets of rules are sorted based on their MCC score. The best set of rules is presented.

Discussion:

We checked the entire diversity within the set of drugs inducing QT prolongation and we have found that they are very highly diverse (**Figure 1**). More than 110 compounds have Tanimoto index of similarity < 0.7. All cardio-toxic

drugs were found to obey Lipinski rule of 5 [19] and Oprea lead-like rule [20]. Drug-like, according to the Lipinski rule of 5, states that orally bio available molecules are more likely to have ≤ 5 H-bond donors, ≤ 10 H-bond acceptors, ≤ 500 molecular weight and ≤ 5 log P. They were derived by an analysis of 2245 drugs. Drug-like molecules should have no more than 1 violation – descriptor's value that is out of the range. However, lead-like, according to Oprea, states that lead molecules are more likely to have ≤ 450 molecular weight, between -3.5 and 4.5 log P, ≤ 4 rings, ≤ 10 non-terminal single bonds, ≤ 5 hydrogen bond donors and ≤ 8 hydrogen bond acceptors. They were derived by an analysis of 96 drugs and leads from which they were derived. We can obviously see that the property criteria for lead-likeness are more stringent (inflexible) than that for the Lipinski rule of 5 for drug-likeness.

The data base available for drugs inducing LQTS was utilized for drawing histograms of five discriminative descriptors, namely the number of nitrogen and oxygen atoms, hydrophobic atoms, single bonds and log P (o/w) (**Figure 2**). These descriptors were employed in this study to construct a useful prediction model that is easy for use to medicinal chemists in drug design. Histograms' construction may help in performing the subsequent division of variable values into two sets: a lower range and an upper range of each descriptor.

Our LQTS rules states that cardio-toxic drugs are more likely to obey Lipinski rule of 5, Oprea lead-like rule as well as having: Log P between -0.5 to 6.5, number of nitrogen atoms between 1 to 5, number of oxygen atoms is up to 4 atoms, number of hydrophobic atoms between 5 to 27, number of single bonds between 15 to 53. Such extracted rules are useful for separating cardio-safe drugs from non-safe ones. A Matthews Correlation Coefficient of 0.6 was attained and nearly 96% of the cardio-toxic drugs (true positives) were successfully identified while only 36% of the cardio-safe drugs (false positives) fall into these ranges. These proposed set of rules could be employed to infer cardio-toxicity or -safety for current and potential drugs. It may have important impact on decision making in the fields of drug development, molecule screening in biological assays, and other applications as well.

Conclusions:

The molecular characters inherent in the investigated drug molecules have shown relevance to cardio-toxicity and -safety. Hence, a set of simple rules has been unravelled and could become a landmark in the predictive ability discriminating between cardio-toxic and - safe drugs. Such set of rules maybe widely employed to determine the cardio-safety potential of molecules, screening molecules for biological assays, as well as for other applications.

Acknowledgments:

We gratefully acknowledge RAND Biotechnologies Ltd for providing us with the database of cardio-toxic and cardio-safe drugs collected from literature.

References:

- [1] B. Darpo et al, *J Clin Pharmacol.*, 46(5):498 (2006) [PMID: 16638733]
- [2] E. Raschi et al., *Pharmacol Res.*, 57(3):181 (2008) [PMID: 18329284]
- [3] A.M. Aronov., *Drug Discov. Today* 10(2):149 (2005) [PMID: 15718164]
- [4] Coi et al, *Eur J Med Chem.* 43(11):2479 (2008) [PMID: 18262683]
- [5] O. Filz et al, *SAR QSAR Environ. Res.* 19(1-2):81 (2008) [PMID: 18311636]
- [6] K. Finlayson et al., *Eur J Pharmacol.* 500(1-3):129 (2004) [PMID: 15464027]
- [7] M. Recanatini et al., *Medicinal Research Reviews* 25:133 (2005) [PMID: 15389727]
- [8] G.A. Gutman et al. *Pharmacol Rev*, 57:473 (2005) [PMID: 16382104]
- [9] R. Brugada et al. *Circulation*, 109:30 (2004) [PMID: 14676148]
- [10] E. Raschi et al., *Pharmacol Rev* 57:181 (2008) [PMID: 18329284]
- [11] M. Recanatini et al., *Med Res Rev*, 25:133 (2005) [PMID: 15389727]
- [12] R.R. Shah, *Pharmacogenomics*, 7(6):889 (2006) [PMID: 16981848]
- [13] B.A. Wible et al, *J Pharmacol Toxicol Methods*, 52(1):136 (2005) [PMID: 15950494]
- [14] A.M. Brown, *Cell Calcium*. 35(6):543 (2004) [PMID: 15110144]
- [15] van Noord et al., *J Clin Psychopharmacol.*, 29(1):9 (2009) [PMID: 19142100]
- [16] Ernest et al., *Ann Pharmacother.*, 42(10):1514 (2008) [PMID: 18728104]
- [17] W. Zareba, *Cardio J.*, 14(6):523 (2007) [PMID: 18651517]
- [18] A.R. Riera et al., *Cardio J.* 15(3):209 (2008) [PMID: 18651412]
- [19] C.A. Lipinski et al., *Adv Drug Deliv Rev.*, 46(1-3):3 (2001) [PMID: 11259830]
- [20] T.I. Oprea et al., *J Mol Graph Model.*, 17(5-6):261 (1999) [PMID: 10840686]

Edited by P. Kanguane

Citation: Falah *et al*, *Bioinformatics* 3(9): 389-393 (2009)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material

Matthews Correlation Coefficient (MCC) equation:

$$MCC = \frac{(PN) - (P_f N_f)}{\sqrt{(N + N_f)(N + P_f)(P + N_f)(P + P_f)}} \quad (1)$$

P and N are the percentages of true positive and true negative predictions while P_f and N_f are the percentages of false positives and false negatives, respectively. True positives are drugs from a database of cardio-safe drugs that are identified as cardio-safe. False positives are cardio-toxic drugs from a cardio-toxic drugs database that are identified as cardio-safe. False negatives are cardio-toxic drugs from a cardio-safe drugs database identified as cardio-toxic, and true negatives are cardio-toxic drugs from a cardio-toxic drugs' database, identified as cardio-toxic. The possible values for MCC range between -1.0 and 1.0 (1.0 for a perfect prediction and -1.0 for a completely erroneous prediction).