

Why are most phospholipidosis inducers also hERG blockers?

Svetoslav Slavov¹ · Iva Stoyanova-Slavova¹ · Shuaizhang Li² · Jinghua Zhao² · Ruili Huang² · Menghang Xia² · Richard Beger¹

Received: 16 February 2017 / Accepted: 18 May 2017 / Published online: 27 May 2017
© Springer-Verlag Berlin Heidelberg (outside the USA) 2017

Abstract Recent reports have noted that a number of compounds that block the human Ether-à-go-go related gene (hERG) ion channel also induce phospholipidosis (PLD). To explore a hypothesis explaining why most PLD inducers are also hERG inhibitors, a modeling approach was undertaken with data sets comprised of 4096 compounds assayed for hERG inhibition and 5490 compounds assayed for PLD induction. To eliminate the chemical domain effect, a filtered data set of 567 compounds tested in quantitative high-throughput screening (qHTS) format for both hERG inhibition and PLD induction was constructed. Partial least squares (PLS) modeling followed by 3D-SDAR mapping of the most frequently occurring bins and projection on to the chemical structure suggested that both adverse effects are driven by similar structural features, namely two aromatic rings and an amino group forming a three-center toxicophore. Non-parametric *U*-tests performed on the original 3D-SDAR bins indicated that the distance between the two aromatic rings is the main factor determining the

differences in activity; at distances of up to about 5.5 Å, a phospholipidotic compound would also inhibit hERG, while at longer distances, a sharp reduction of the PLD-inducing potential leaves only a well-pronounced hERG blocking effect. The hERG activity itself diminishes after the distance between the centroids of the two aromatic rings exceeds 12.5 Å. Further comparison of the two toxicophores revealed that the almost identical aromatic rings to amino group distances play no significant role in distinguishing between PLD and hERG activity. The hypothesis that the PLD toxicophore appears to be a subset of the hERG toxicophore explains why about 80% of all phospholipidotic chemicals (the remaining 20% are thought to act via a different mechanism) also inhibit the hERG ion channel. These models were further validated in large-scale qHTS assays testing 1085 chemicals for their PLD-inducing potential and 1570 compounds for hERG inhibition. After removal of the modeling and experimental inconclusive compounds, the area under the receiver-operating characteristic (ROC) curve was 0.92 for the PLD model and 0.87 for the hERG model. Due to the exceptional ability of these models to recognize safe compounds (negative predictive values of 0.99 for PLD and 0.94 for hERG were achieved), their use in regulatory settings might be particularly useful.

The views presented in this article are those of the authors and do not necessarily reflect those of the US Food and Drug Administration. No official endorsement is intended nor should be inferred.

Electronic supplementary material The online version of this article (doi:10.1007/s00204-017-1995-9) contains supplementary material, which is available to authorized users.

✉ Richard Beger
Richard.Beger@fda.hhs.gov

¹ Division of Systems Biology, National Center for Toxicological Research, 3900 NCTR Road, Jefferson, AR 72079, USA

² National Center for Advancing Translational Sciences, Bethesda, MD, USA

Keywords Human Ether-à-go-go related gene · Phospholipidosis · 3D-SDAR · Toxicophore · Molecular modeling

Introduction

Both drug-induced phospholipidosis (PLD) and human Ether-à-go-go related gene (hERG) potassium channel

blocking are undesired side effects of drugs. While it is still unclear if the phospholipidosis causes any adverse health effects on humans (Reasor and Kacew 2001), blocking of the hERG channel has been associated with prolonged QT intervals, which may further degenerate to Torsades de pointes and in severe cases to sudden cardiac death (Witchel 2011). PLD induction, hERG blocking, and the link between PLD and hERG (Sun et al. 2013) are still a major concern to the pharmaceutical industry during the preclinical testing phase of drug candidates. The increased interest towards these two adverse events has led to numerous efforts to model the PLD-inducing and hERG blocking potentials of chemicals. Including our own (Stoyanova-Slavova et al. 2017), more than 70 models of hERG blocking (Villoutreix and Taboureau 2015) have been published to date. Unlike hERG, most PLD models (Bauch et al. 2015) were published after 2008 when a large curated phospholipidosis database compiled at FDA became publicly available (Kruhlak et al. 2008).

Despite the substantial amount of PLD and hERG models reported in the literature, only a few attempts to associate the spatial configuration of specific substructural units with activity (Cavalli et al. 2002; Goracci et al. 2015; Slavov et al. 2014; Stoyanova-Slavova et al. 2017) have been published. Among these, our earlier work on hERG (Stoyanova-Slavova et al. 2017) was the first one to emphasize the similarity between the hERG and PLD toxicophores. In both cases, three-center toxicophores composed of two aromatic rings and an amino group were found to be present in the structures of most phospholipidotic and hERG inhibiting drugs. Furthermore, we hypothesized that the gross-similarity between the two toxicophores is responsible for the fact that many hERG channel blockers are also PLD inducers (Sun et al. 2013). Our models indicated that the distance between the two aromatic rings in the molecules of phospholipidotic compounds varies in a narrow range, between 4 and 5 Å; in the structures of hERG blockers, this distance spans a much wider range from about 4.5–11.5 Å. In both cases, the aromatic ring-to-amino-group distances were found to be approximately the same, and thus unlikely to play a role in distinguishing between hERG and PLD active compounds. As a result of these observations, we put forward a hypothesis that the distance between the two aromatic rings is the defining factor for the activity mode—at shorter distances, a phospholipidotic compound would also inhibit the hERG ion channel, while at longer distances, a sharp reduction of the PLD-inducing potential would leave only a well-pronounced hERG blocking effect. The Venn diagram reported by Sun et al. (2013) appears to be in agreement with such a conjecture, indicating that about 80% of all PLD inducers would also block hERG, a likely consequence of the fact that the PLD toxicophore appears to be a subset of the hERG toxicophore. However,

since the PLD and hERG data sets used to derive the above hypothesis did not cover identical chemical domains (only alosetron, clozapine, haloperidol, and thioridazine were common to both), it is unclear whether such an interpretation reflects the “true” nature of the underlying biochemical phenomena or is a mere statistical artifact, a function of the structural diversity and dissimilarity between the two data sets.

To further elucidate the role of the spatial configuration of the toxicophore centers, a set of compounds tested for both their PLD-inducing and hERG blocking potentials was compiled—the uniformity of this data set allowed exploration of the underlying structural factors by cancelling out the effects caused by differences in the covered chemical domains. In addition to the above-defined goal, this work intended to achieve the following:

1. develop reliable, OECD compliant models for prediction of phospholipidosis induction and hERG inhibition;
2. derive toxicophores, whose presence is associated with PLD and/or hERG activity;
3. determine how the PLD and hERG activity depends on the toxicophores' geometry;
4. test the validity of our earlier hypothesis that the distance between the two aromatic rings determines the difference in biological activity;
5. validate experimentally and compare the predictive performance of both models.

Data set

HERG and PLD assays in a qHTS format

A slightly modified version of the qHTS assay using human U-2 OS (osteosarcoma) transduced cells (Titus et al. 2009) was used to evaluate the hERG inhibition of a total of 4095 chemicals (Sun et al. 2013). After culturing in Dulbecco's Modified Eagle Medium (Invitrogen, Carlsbad, CA, USA) with Glutamax containing 10% fetal bovine serum (HyClone, Logan, UT, USA), 1% of Non-Essential Amino Acids (Invitrogen), and 50 U/mL penicillin/50 µg/mL streptomycin (Invitrogen), the U-2 OS cells were transduced using a BacMan-hERG construct (Montana Molecular, Bozeman, MT, USA) and subjected to a thallium influx assay (Xia et al. 2011). Data analysis performed as previously described by Titus et al. (2009) followed by classification based on the type of the observed concentration–response curves (Wang et al. 2010) assigned each compound to one of the following three categories: active, inactive, or inconclusive.

A qHTS assay described by Shahane et al. (2014) measuring the fluorescence intensity of LipidTOX red dye (Molecular Probes, H34351) labeled phospholipid accumulations in Human HepG2 (Hepatocellular carcinoma) cells (ATCC, Manassas, VA, USA) was used to quantify the PLD-inducing potential of 5490 chemicals. Amiodarone, a well-known PLD inducer, and DMSO were used as positive and negative controls, respectively. The fluorescence intensities (595 nm excitation, 615 nm emission for LipidTox red) were measured using an ImageXpress Micro Widefield High Content Screening System (Molecular Devices, Sunnyvale, CA, USA) with a 20X Plan Fluor objective. Data analysis was performed as previously described by Wang and Huang (2016). The same three categories (active, inactive, or inconclusive) were defined.

Data set design

Among the 4095 chemicals assayed for their hERG inhibition and 5490 chemicals tested for their phospholipidotic potential, a total of 2456 chemicals were common to both sets (see Table 1). The inability to formulate a rational rule for class assignment (as either active or inactive) led to the removal of 328 hERG and 122 PLD inconclusive compounds. Furthermore, all duplicates, mol files containing multiple structures, metal, ammonium, or sulfonate ions were also excluded.

As can be seen from Table 1, the original data set (this was also true for the curated set) contained a disproportionately large number of inactive chemicals. Since models based on heavily unbalanced data sets tend to be biased in the direction of the majority class and thus may be unable to capture well structural trends associated with the minority class (He and Garcia 2009), undersampling strategies reducing the number of PLD and hERG inactive compounds were applied. Since earlier studies indicated that the structures of most PLD and hERG actives contain at least one aromatic ring and a nitrogen atom (Slavov et al. 2014; Stoyanova-Slavova et al. 2017), therefore, the task of building binary classification models for which the chemicals in the active class are substantially dissimilar from these in the inactive class was abandoned as being trivial. For example, building a classification model for which the active class is comprised of aromatic amines (typical PLD and hERG actives), whereas the inactive class contains primarily aliphatic, non-nitrogen containing chemicals constitutes a trivial modeling exercise with a little added value. Hence, a robust modeling data set comprised of either active or inactive nitrogen containing aromatic compounds was constructed. Thus, the generated PLD and hERG 3D-SDAR models should be able to capture small variations in the chemical structure leading to substantial changes in biological activity. The high degree of structural

Table 1 hERG and PLD compounds by class, shown for the reduced set of 2456 chemicals for which PLD and hERG data were available

	PLD class		
	Inactive	Inconclusive	Active
hERG class	1741	51 ^a	41
Inactive			
Inconclusive	270 ^a	27 ^a	31 ^a
Active	159	44 ^a	92

^a All chemicals with inconclusive data were excluded from further analysis

similarity between the chemicals in the active and inactive classes can be regarded as a prerequisite for establishing a sound structure–activity relationship built on solid chemical and biological foundations.

The removal of all aliphatic chemicals while retaining all nitrogen containing aromatic compounds still resulted in a heavily unbalanced data set with many more inactive than active chemicals. Focused on small molecules, further reduction of the number of inactive samples was achieved by filtering out all chemicals with a molecular weight exceeding 300 AU. To retain most of the relatively few samples in the active class, only compounds for which the prediction of the nitrogen atom(s) chemical shifts failed as well as a small subset of 28 steroid derivatives (due to the different mechanism by which they elicit their PLD effect) were removed. The final data set consisted of 567 chemicals common to both PLD and hERG (listed in supplementary information spreadsheet). Of these two, the hERG data set was somewhat better balanced: 234 compounds were hERG active (further denoted as hERG+), while the remaining 333 chemicals were hERG inactive (or hERG–). On the other hand, the ratio between the active (108) and inactive (459) samples in PLD data set was approximately 1:4.

Methodology

3D-SDAR fingerprint construction and descriptor generation

The optimized geometries of all 567 compounds and their corresponding ¹³C and ¹⁵N chemical shifts simulated by ACD/NMR Predictor version 12.0 (Advanced Chemistry Development, Toronto, Canada 2011) were used to generate unique 3D-SDAR molecular fingerprints as described previously (Slavov et al. 2014; Stoyanova-Slavova et al. 2017). In brief, a 3D-SDAR fingerprint is composed of $n(n-1)/2$ fingerprint elements (where n is the number of C and N atoms in a molecule), each of which has X and Y coordinates determined by the NMR chemical shifts δ_{Ai}

and δ_{A_j} of a pair of atoms ($A_i, A_j \equiv C, N$) and a Z coordinate corresponding to the through-space distance (r_{ij}) between them. This concept is illustrated in Fig. 1, which shows the structure and the ^{13}C and ^{15}N NMR chemical shifts of aniline (Fig. 1a) as well as its corresponding 3D-SDAR fingerprint (Fig. 1b).

These fingerprints were then tessellated by regular grids and the count of fingerprint elements within the boundaries of each bin was stored in a 3D-SDAR descriptor matrix. Earlier models built for non-overlapping PLD (Slavov et al. 2014) and hERG (Stoyanova-Slavova et al. 2017) data sets focused on a more detailed exploration of the 3D-SDAR parametric space suggested optimal models using the following grid parameters: 6C30N1Å or 10C25N0.5Å for hERG and 8C20N1Å or 10C25N1Å for PLD. However, models using tessellations different from the above were found to perform similarly with only a minor drop of about 5% in terms of overall accuracy, sensitivity, and specificity. To be able to compare directly the performance

characteristics, the generated 3D-SDAR maps, and their corresponding toxicophores, the present models utilized a fixed grid size of 10 ppm \times 10 ppm \times 0.5Å grid for the C–C region, 10 ppm \times 25 ppm \times 0.5Å grid for the C–N region, and 25 ppm \times 25 ppm \times 0.5Å grid for the N–N region (or 10C25N0.5Å in short notation).

Modeling algorithm

A bootstrap aggregation (bagging)-based PLS algorithm (shown in Fig. 2) was used to establish a consistent relationship between the 3D-SDAR descriptors and the PLD and hERG class membership. Earlier computational experiments (Slavov et al. 2013) using reasonably sized data sets demonstrated that most performance estimators tend to reach a plateau at about 100 randomization cycles. Hence, the employed PLS algorithm was set to perform 100 random splits partitioning the modeling set into training (80% of the total) and hold-out test (20% of the total) sets. Models with up to ten

Fig. 1 3D-SDAR fingerprint (b) of aniline (a) constructed using the atom-to-atom through-space distances and the atoms' corresponding ^{13}C (shown in black) and ^{15}N (shown in blue) NMR chemical shifts. For simplicity, the $\delta^{15}\text{N}$ was shifted downfield by 1000 ppm

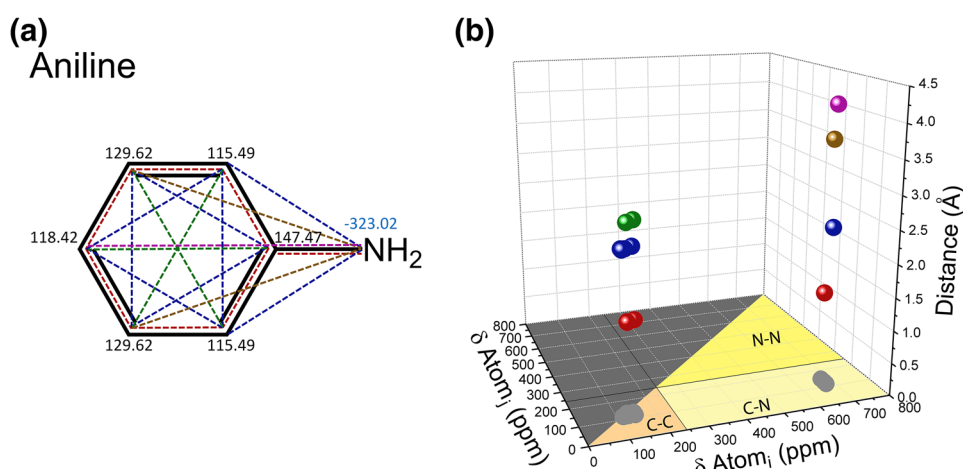
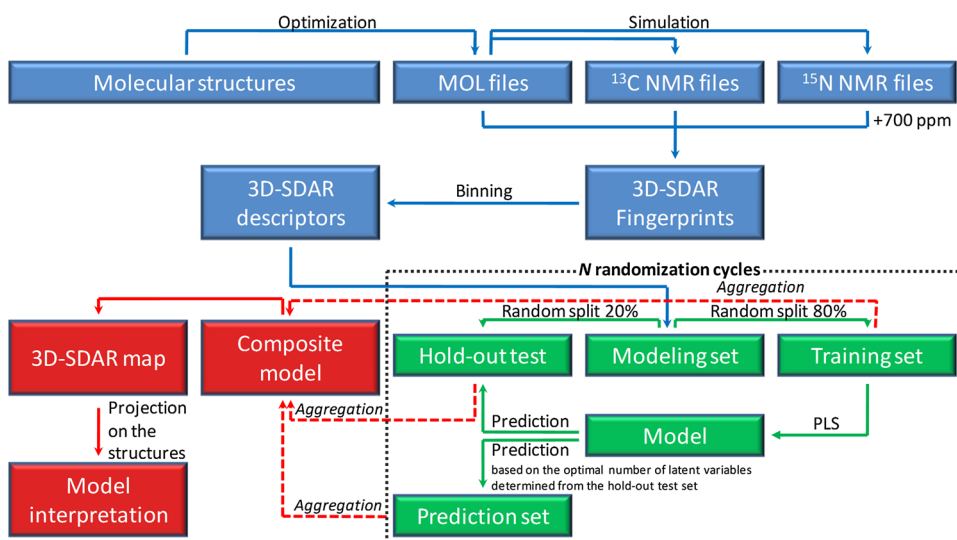


Fig. 2 Flowchart of the 3D-SDAR modeling process (adapted from Ref. [4]). A bagging-like PLS (partial least squares) algorithm was used to build a composite model, which was used further to elucidate the structural features associated with PLD induction and hERG blocking (color figure online)



latent variables (LVs) were generated. At the end, the aggregated predicted values for each individual compound were averaged (that was done separately for the training and hold-out sets) and cut-off values equal to the ratio between the positive samples and the total number of compounds in each of the two (PLD and hERG) data sets was used for class assignment. The optimal number of LVs was determined from the plateau on the LVs vs classification accuracy plot for the hold-out test set. These two models were later used to predict the PLD and hERG activity of compounds from the Tox21 program. Due to the specifics of the bagging-like PLS approach, each compound from the modeling set had two distinct average predicted values—functions of its random assignment to either the hold-out test set or the training set. Since the training set statistical parameters are inconsistent predictors of behavior of the models in situations in which the activity of novel/untested compounds needs to be predicted, our further discussion will be focused only on the hold-out test set performance characteristics and the models' true external predictive power.

Interpretability

The final stage of the modeling phase involved deciphering of the structure–activity relationship using 3D-SDAR maps of the most frequently occurring bins (Slavov et al. 2014). These bins are usually clustered together in specific regions of the 3D-SDAR space; their location on the *XY*-plane allows for an identification of the structural features associated with activity, while their position on the *Z*-axis determines the distances between these features. Since there is only a limited number of interactions and orientations by which a compound can activate a receptor and exert its activity and a multitude of factors preventing such interactions leading to inactivity, for simplicity, these maps were generated only for the highest positively weighted bins. These bins were further projected on the standard coordinate space to determine the spatial configuration of the toxicophore centers. A non-parametric Mann–Whitney *U*-test exploring the dependence between bin occupancy and activity at fixed chemical shifts (on the *XY*-plane) and varying distances (on the *Z*-axis) was used to determine the most optimal distances between the toxicophore centers specific to strong hERG blockers and PLD inducers.

Optimal prediction space and applicability domain determination

The use of 3D-SDAR fingerprints allows for a relatively straightforward determination of the applicability domains (AD) of models. Whether a new/untested compound belongs to the AD of a 3D-SDAR model is determined by the similarity of its fingerprint to the fingerprint

of a compound present in the training set. Within the framework of 3D-SDAR, the similarity between two molecular fingerprints is calculated from their linearized representations (equivalent to the rows in a 3D-SDAR matrix) using the generalized Tanimoto similarity (Bajusz et al. 2015) formula $T(A, B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B}$, in which *A* and *B* are vectors of bin occupancies. Instead of relying on arbitrarily defined thresholds, 3D-SDAR uses an objective procedure to determine if a new/untested compound belongs to the AD of a model. These thresholds were determined from the maximum pairwise similarities (T_{\max}) calculated for all possible pairs of chemicals in the training set: the average and the standard deviation of these T_{\max} values can be used to define four AD regions with predictions ranging from excellent to poor. Compounds characterized by $T < T_{\text{mean. trn.}} - \text{SD}_T$ are considered outside of the AD. The application of this approach will be illustrated in the “Experimental validation” section.

Results and discussion

Modeling performance

As described in the “modeling algorithm” section, applying cutoffs of 0.190 in the case of PLD and 0.413 in the case of hERG allowed a conversion of the generated continuous predicted values back to categorical class assignments. While in the case of hERG, the accuracy, sensitivity, and specificity leveled off after 2 LVs, in the case of PLD, the overall accuracy plateaued at 6 LVs (0.799), whereas the sensitivity achieved its highest value at 2 LVs (0.889). Since parsimonious models (those with fewer LVs) are generally preferred (Cherkasov et al. 2014), in

Table 2 Statistical performance of the hERG and PLD models for the hold-out test set

Performance/set	PLD	HERG
Accuracy	0.785	0.887
Sensitivity	0.889	0.829
Specificity	0.760	0.928
Negative predictive values	0.967	0.885
Positive predictive values	0.466	0.890
Predicted positive	206	218
Predicted negative	361	349
True positive	96	194
True negative	349	309
False positive	110	24
False negative	12	40

both cases, 2 LVs were considered optimal and used further. Table 2 summarizes the statistical parameters of our hERG and PLD models using 10C25N0.5Å bin size and 2 LVs.

It is interesting to note that although the present models were based on a larger and more diverse data set, their performance exceeded that of our earlier attempts to model PLD and hERG (Slavov et al. 2014; Stoyanova-Slavova et al. 2017). Compared against each other, the present hERG model is characterized by a somewhat lower sensitivity, while its overall accuracy is significantly higher than that of the PLD model. Furthermore, due to very few positive samples in the PLD data set which affected negatively the ability of the PLS algorithm to capture structural trends associated with a significant phospholipidotic potential, the model's positive predicted value is almost twice as low as that of the hERG model.

Structural interpretation

The atom level resolution of 3D-SDAR combined with the linearity of PLS allows for a relatively straightforward identification of toxicophores or pharmacophores in the structures of bioactive chemicals. However, whereas a standard PLS algorithm could utilize directly the weights of individual bins as a measure of their statistical importance and provide an interpretation based on that, the bagging-like algorithm driving 3D-SDAR requires a different approach. As described in detail in Stoyanova-Slavova et al. (2017), the simplest alternative is to extract a predefined number of highly ranked, positively weighted bins from each latent variable for all individual models forming the final composite model and use their frequency of occurrence instead of their individual PLS weights. In general,

if a given bin (3D-SDAR descriptor) appears to be statistically significant (has a high PLS weight) and occurs frequently, then it is more likely that it depicts a structural feature essential for activity. The rationale behind this choice was that: (1) as a grid-based approach 3D-SDAR generates thousands of descriptors and (2) within smaller groups of descriptors ranked according to their PLS weights, the weights vary insignificantly (i.e. these descriptors can be regarded as contributing equally). In other words, to decode the structure–activity relationships established by the 10C25N0.5Å models, the top 10 positively weighted bins for each of the two LVs, for all 100 randomized models, were extracted and their frequencies of occurrence were calculated and mapped on the 3D-SDAR space. Figure 3 shows the 3D-SDAR maps corresponding to the PLD and hERG models, whose statistical parameters are carried out in Table 2.

As can be seen from Fig. 3 the PLD and hERG 3D-SDAR maps are substantially similar sharing five distinct clusters at similar/identical locations. As explained earlier, the positions of these bins on the XY-plane identify uniquely the atom-pair types and their Z coordinate describes the through-space distance between the two atoms. The bins within the black contour that are close to the origin describe the spatial relationship between two carbon atoms connected directly to a nitrogen atom. The cluster of bins within the blue contour represents aromatic carbons at varying proximity (4Å to 9Å) to carbons that are immediate neighbors of a nitrogen atom. The red contour encloses bins describing aromatic carbons members of the same (at short distances) or different (distances larger than 3 Å) rings. The carbon–nitrogen bins within the green contour depict an amino group, which is generally 6.5 to 8.5 Å apart from the carbons of an aromatic ring system.

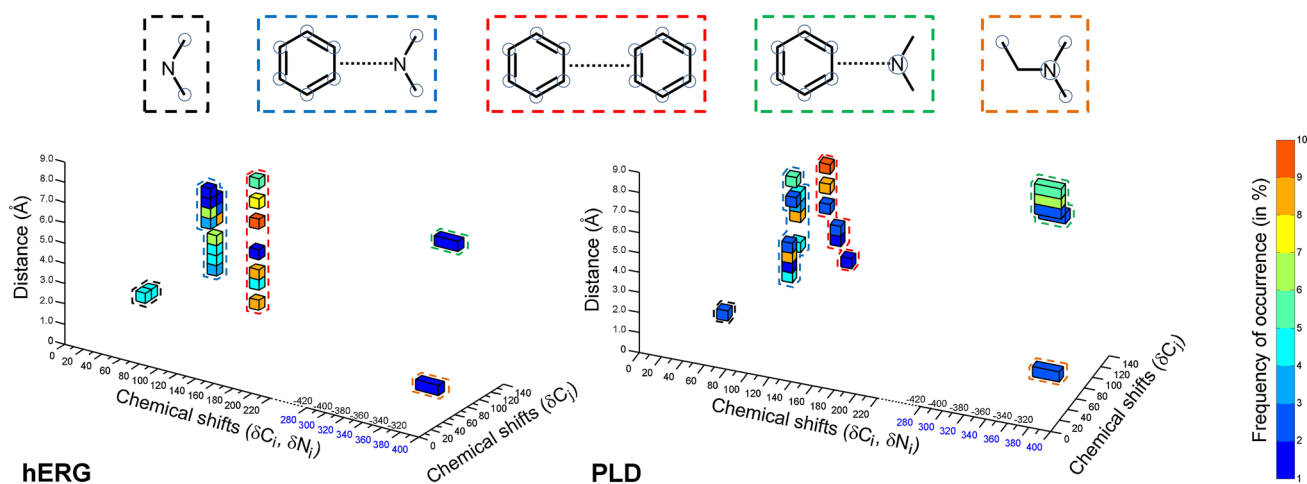


Fig. 3 3D-SDAR map of the most frequently occurring C–C and C–N bins. For convenience, the $\delta^{15}\text{N}$ are shifted by 700 ppm past the range of $\delta^{13}\text{C}$. The original negative $\delta^{15}\text{N}$ values are shown in black

The single bins in the orange contours describe the spatial relationship between a nitrogen atom and its first- or second-order aliphatic carbon neighbors. As evident from the 3D-SDAR maps and their associated fragments, the presence of two aromatic rings and an amino group are the characteristic features of both PLD inducers and hERG blockers.

Toxicophore comparison

Projection of the most frequently occurring bins on to the molecular structures elucidated further the characteristics of the PLD and hERG toxicophores. As suggested by the 3D-SDAR maps shown in Fig. 3, the molecules of many PLD inducers and hERG blockers contained three toxicophore centers: two hydrophobic (aromatic rings) and one hydrophilic (amino group). However, in several instances, it was observed that even a single hydrophobic center is sufficient for activity, thus indicating the optional inclusion of the second aromatic rings (shown using dashed lines). A CoMFA analysis reported by Cavalli et al. (2002) reached similar conclusions, describing a hERG toxicophore comprised of one to three hydrophobic centers (aromatic moieties) and an amino group.

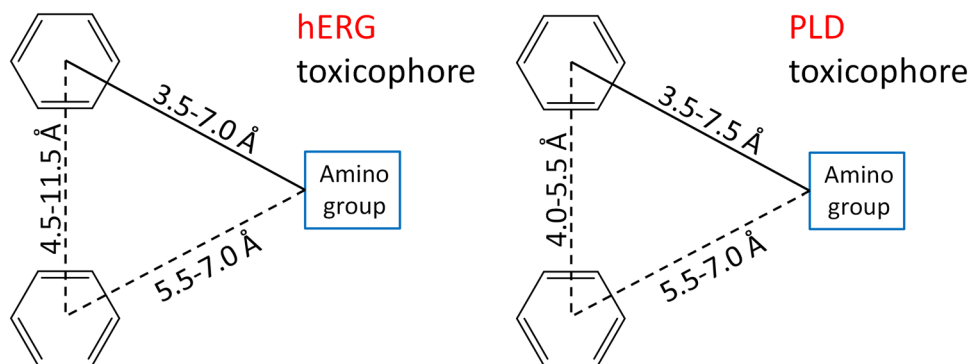
Exploring further the information encoded along the Z-axis of the 3D-SDAR space allowed the determination of a range of optimal distances between the three toxicophore centers. These distances are shown along the connecting solid and dashed lines in Fig. 4. A comparison between the hERG and PLD toxicophores revealed that the distances between the aromatic rings and the amino groups are almost identical and, therefore, cannot be considered as factors contributing to the observed differences in activity. However, it appeared that the distance between the two aromatic rings in the molecules of hERG blockers is less constrained compared to that found in the structures of PLD inducers—in other words, the PLD toxicophore is a subset of the hERG toxicophore. An important corollary derived from these results is that most phospholipidic compounds will also block the hERG channel,

whereas hERG blockers in which the aromatic rings are farther than 5.5 Å apart would not induce phospholipidosis. This hypothesis is supported by data from large-scale qHTS hERG inhibition and PLD-inducing potential assays (Sun et al. 2013), which reported that about 80% of all phospholipidic compounds are also hERG blockers, with the remaining 20% (all steroid derivatives) likely acting via a different mechanism. On the other hand, only about 40% of the hERG inhibitors were able to induce phospholipidosis.

This hypothesis was further tested by performing a Mann–Whitney *U*-test comparing the bin occupancy of the active versus inactive chemicals at fixed chemical shifts in the *XY*-plane and varying distances on the *Z*-axis. In other words, the association of the bin occupancy with the probability of a compound being active or inactive as a function of the distance between the toxicophore centers was explored. This method allowed for a direct comparison of the most favorable spatial configurations of the hERG and PLD toxicophore centers. However, since it is based on the original through space atom-to-atom distances, this approach cannot distinguish between the two aromatic rings to amino-group distances (characteristic for each toxicophore), as the carbon atoms of both rings occupy the same bins.

The *p* values for two groups of bins were calculated: (1) $X = 120\text{--}130$ ppm, $Y = -361$ ppm to -336 ppm and $Z = 1\text{--}17$ Å (step 0.5 Å) corresponding to the distances between aromatic ring carbons and a nitrogen atom and (2) $X = 120\text{--}130$ ppm, $Y = 120\text{--}130$ ppm and $Z = 1\text{--}20$ Å (step 0.5 Å) corresponding to the distances between aromatic carbons, members of the same (distances <3 Å) or two different ring systems (distances >3 Å). Due to the use of the original atom-to-atom distances, the distances shown in Fig. 5a and b differ slightly from those based on centroids (Fig. 4). In Fig. 5a and b, low logarithmic *p* values correspond to bins, which are predominantly occupied by fingerprint elements of either PLD or hERG actives, but are rarely hit by fingerprint elements belonging to inactive chemicals.

Fig. 4 hERG and PLD toxicophores derived from projection of the most frequently occurring bins on the chemical structures. The distances between the toxicophore centers are given to their corresponding centroids



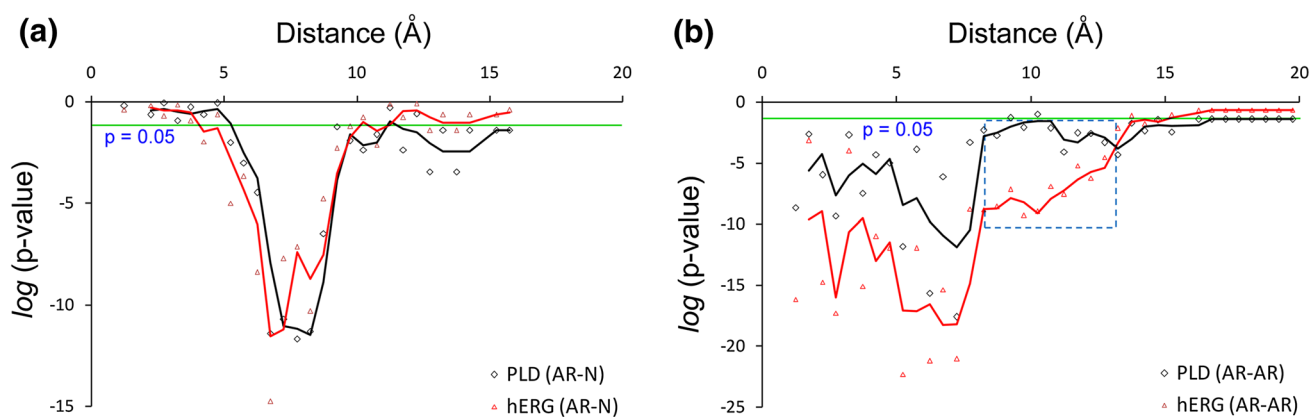


Fig. 5 Moving average $\log(p)$ value curves for the distances between: **a** aromatic ring carbon atoms and the amino-group nitrogen corresponding to the 120–130 ppm, –361 to –336 ppm, $Z = 1 \dots 17$

b carbons from two different aromatic rings (120–130 ppm, 120–130 ppm, $Z = 1 \dots 20$ Å)

As evident from the logarithmic scale used in Fig. 5a and b, most bins are highly specific (p values of 10^{-10} and lower) and are occupied almost exclusively by the fingerprint elements of active chemicals. The almost identical curves shown in Fig. 5a demonstrate clearly that the aromatic-to-amino-group distances are non-specific and cannot explain the observed differences in activity between the PLD and hERG chemicals. Similar to the toxicophores shown in Figs. 4, 5b indicates the critical importance of the distance between the carbon atoms belonging to two different aromatic rings; at shorter distances of up to about 7 Å, the chemicals are both PLD inducers and hERG blockers, while at longer distances (shown in the blue contour in Fig. 5b), they lose their phospholipidotic potential, but retain their hERG blocking ability. The hERG blocking potential decreases sharply after about 12.5 Å (Fig. 5b).

Prediction of compounds for experimental validation

Similar to the model building phase, the true predictive power of the PLD and hERG models was tested by carefully avoiding the most generic (and with a somewhat predetermined outcome) forms of experimental validation: i.e., combining for example chemicals containing the simplest toxicophore (an aromatic ring and an amino group) that have a high probability of being active with aliphatic chemicals which would likely be inactive. Hence, to rigorously test the proposed models, the Tox21 10 K chemical library (NCATS 2016; PubChem 2013) was screened for compounds containing at least one aromatic ring and a nitrogen atom; applying the two models these were classified into active and inactive chemicals. This specific choice made our task increasingly more difficult: since all compounds, both active and inactive, contain at least two toxicophore centers, the models should be able to make an

accurate distinction based on small variations in the chemical structures (substituents) and the distances between the toxicophore centers. Due to unavailability or lack of sufficient amounts for testing, the filtered list of compounds was reduced to 1823 chemicals whose hERG inhibition was to be predicted and 1167 chemicals (a subset of the above list of 1823 compounds) to be classified according to their phospholipidotic potential. A total of 304 of these were predicted as hERG+, whereas 247 were predicted as PLD+. A summary of our predictions for all tested compounds as well as their chemical names and experimental activities are given in Table S11.

Optimal prediction space and domains of applicability

Within the framework of 3D-SDAR, the Tanimoto similarity as defined in the “Optimal prediction space and applicability domain determination” section is used to determine several regions of reliability/confidence in prediction. These regions are defined on the basis of the distribution of the T_{\max} values within the training set. These T_{\max} values are calculated from the maximum degree of similarity between any one chemical from the training set and the remaining 566 chemicals. The histogram of their distribution is shown in Fig. 6 using darker gray-shaded bars. If a new/untested compound has an analogue in the training set to which it is similar with T exceeding $T_{\text{tr.set mean}} + \sigma_T$; its prediction will be highly reliable (see the region of “excellent” predictivity in Fig. 6), in the range between $T_{\text{tr.set mean}}$ and $T_{\text{tr.set mean}} + \sigma_T$; the predictions will be good, below that region ($T_{\text{tr.set mean}} - \sigma_T$ and $T_{\text{tr.set mean}}$); the predictive accuracy will be fair, and if a compound has no close analogue in the training set ($T_{\text{tr.set mean}} - \sigma_T$), it will be considered outside of the applicability domain and, respectively, predicted poorly. In Fig. 6, the lighter bars represent the

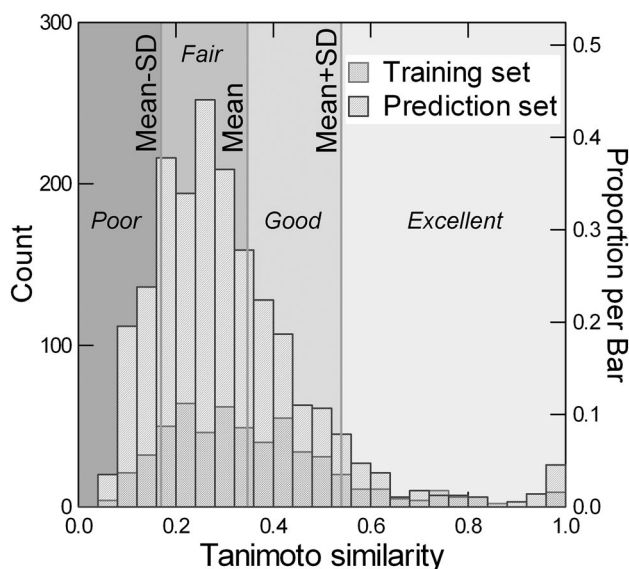


Fig. 6 Histograms of the Tanimoto similarity distribution for: (1) the training set (*darker gray bars*) and (2) for the prediction set (*lighter gray bars*)

distribution of the T_{\max} values for the 1823 predicted chemicals—approximately 1/5 of them lay outside of the applicability domain and their class assignment is unreliable.

Experimental validation results

Predictions from our 3D-SDAR models were submitted to the National Center for Advancing Translational Sciences (NCATS) which upon availability tested 1570 and 1085 chemicals for their hERG blocking and PLD-inducing

potentials, respectively. After removing the experimental and predicted ($\pm 10\%$ around the cut-off values) inconclusive compounds, the true predictive performance of the 3D-SDAR models based on 1383 hERG and 1012 PLD assayed chemicals was evaluated. Since the bagging-like PLS algorithm produces a quantitative output the receiver-operating characteristic (ROC) curves and their corresponding area under the curve (AUC) values were calculated and are shown in Fig. 7 with a blue line. As can be seen from Fig. 7, both curves manifest similar behavior with an AUC close to 0.90. The removal of the out of AD chemicals did not improve these results significantly (blue curves).

A comparison of the accuracy, sensitivity, and specificity of the PLD predictions shown in Table 3 to their corresponding values for the hold-out test set from Table 2, demonstrated excellent transferability of internal-to-external predictivity. Although in terms of overall accuracy, the hERG model outperformed the PLD model, its sensitivity was lower. However, a comparison of the positive predictive values indicated that the hERG model recognized correctly twice as large of a proportion of the positive samples. Furthermore, both PLD and hERG models were characterized by exceptionally high negative predictive values, which in case of phospholipidosis reached 0.99. Due to their ability to recognize correctly (with just a few exceptions) most safe compounds, both models should be particularly valuable for use in regulatory settings.

Further examination of the dependence of the accuracy, sensitivity, specificity, and negative and positive predicted values (see Fig. 8) from the above-defined four confidence regions (poor, fair, good, and excellent)

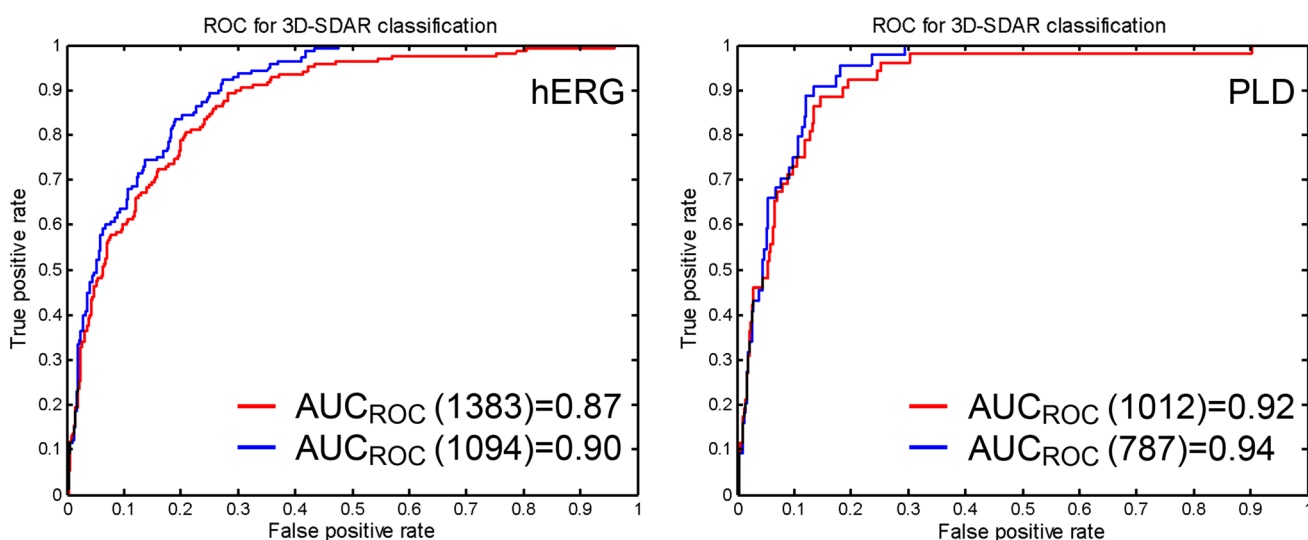


Fig. 7 ROC curves (shown in *red*) for the predictions based on the PLD and hERG 3D-SDAR models. The recalculated ROC curves after removing the out of AD compounds are shown in *blue* (color figure online)

Table 3 Number of compounds in the predicted, tested, and validation sets

Sets/performance characteristics	PLD	HERG
Predicted	1167	1823
Tested by NCATS	1085	1570
Inconclusives removed	1012	1383
Accuracy	0.844	0.863
Sensitivity	0.885	0.588
Specificity	0.842	0.902
Negative predictive values	0.993	0.940
Positive predictive values	0.232	0.457

ACC, SN, and SP reported for the validation set

revealed several trends in the data. While the overall accuracy, specificity, and negative predictive values improved only slightly, the sensitivity and the positive predictive values improved tremendously moving from the region of poor predictions to the region of excellent predictions. This observation further emphasizes the fact that all active chemicals are characterized by highly selective functional groups/substituents positioned at specific locations and with well-defined spatial orientations—hence, a high degree of structural similarity (i.e. fingerprint similarity) between a new untested chemical and a chemical from the training set is required for an accurate prediction of activity. However, in case of hERG- and PLD-chemical fingerprints, a similarity lower than $T_{\text{tr.set mean}} - \sigma_T$ is still sufficient for a correct classification. The high structural specificity required for PLD (as well as for hERG) activity suggests that the phospholipidosis is likely a receptor driven/mediated phenomena.

Due to the simplicity of the derived toxicophores, a simple visual inspection can be used as a quick filter to decide whether a new untested compound should be scrutinized and submitted for in silico modeling and further laboratory testing.

Conclusions

The 3D-SDAR modeling performed on data sets of overlapping chemicals tested for their PLD-inducing and hERG blocking potentials resulted in models with an excellent external predictive power as demonstrated by the qHTS validation assays. There are several corollaries that follow naturally from the analysis carried out in this work:

1. Many PLD and hERG active chemicals share common structural features—two aromatic rings and an amino group forming a three-center toxicophore;
2. the PLD and hERG toxicophores are characterized by identical aromatic ring-to-amino-group distances. However, in the molecules of hERG blockers, the distance between two aromatic rings varies to a much greater extent;
3. compared to the hERG toxicophore, the PLD toxicophore is geometrically more constrained and appears to be a subset of hERG. Hence, most PLD inducers would also block the hERG ion channel, whereas hERG inhibitors with larger aromatic-to-aromatic ring distances would not induce phospholipidosis.
4. the exceptionally high negative predictive values of both models make them potential candidates for use in regulation;
5. the reason for the apparent similarity between the PLD toxicophore and the structures of hERG inhibi-

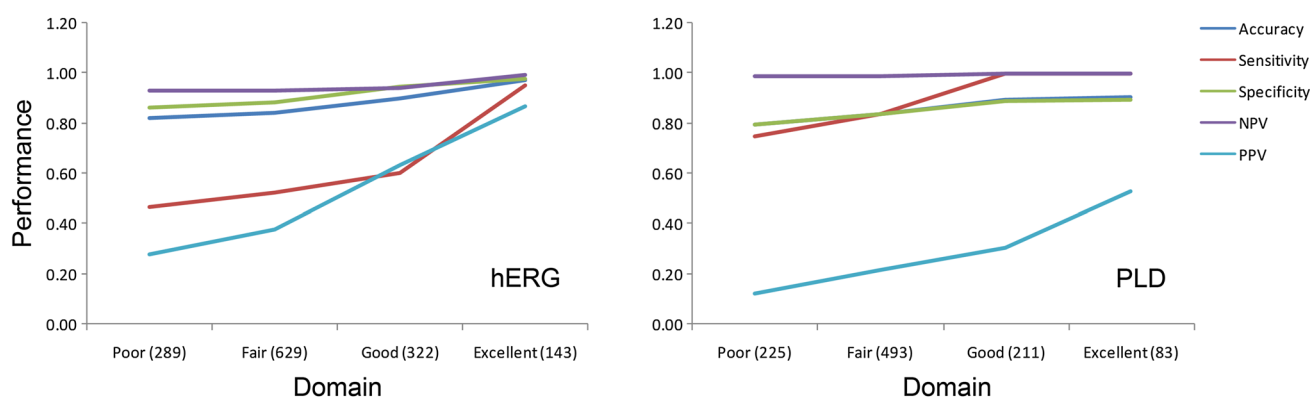


Fig. 8 Dependence of the hERG and PLD models' predictive performance characteristics from the degree of similarity to the training set chemicals

tors remains unknown at this time and requires further investigation.

This work also demonstrated the capability of 3D-SDAR to provide structural interpretation of the models in terms of toxicophores or pharmacophores. Due to its atom level resolution 3D-SDAR can also determine a set of optimal distances between the toxicophore/pharmacophore centers, which makes it a valuable tool with application in the fields of computational toxicology and molecular design.

Acknowledgements This work was partially supported by the U.S. Environmental Protection Agency (Interagency Agreement #Y3-HG-7026-03) and the interagency agreement IAG #NTR 12003 from the National Institute of Environmental Health Sciences/Division of the National Toxicology Program to the National Center for Advancing Translational Sciences, National Institutes of Health. We would like to thank Sampada Shahane for her technical support.

References

- ACD/NMR Predictor Release 12.00, version 12.5; advanced chemistry development. Toronto, Canada, 2011
- Bajusz D, Rácz A, Héberger K (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* 7:20–33
- Bauch C, Bevan S, Woodhouse H, Dilworth C, Walker P (2015) Predicting in vivo phospholipidosis-inducing potential of drugs by a combined high content screening and in silico modelling approach. *Toxicol In Vitro* 29:621–630
- Cavalli A, Poluzzi E, De Ponti F, Recanatini M (2002) Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K⁺ channel blockers. *J Med Chem* 45:3844–3853
- Cherkasov A, Muratov EN, Fourches D, Varnek A, Baskin II, Cronin M, Dearden J, Gramatica P, Martin YC, Todeschini R, Consonni V, Kuz'min VE, Cramer R, Benigni R, Yang C, Rathman J, Terfloth L, Gasteiger J, Richard A, Tropsha A (2014) QSAR modeling: where have you been? Where are you going to? *J Med Chem* 57:4977–5010
- Goracci L, Buratta S, Urbanelli L, Ferrara G, Di Guida R, Emiliani C, Cross S (2015) Evaluating the risk of phospholipidosis using a new multidisciplinary pipeline approach. *Eur J Med Chem* 92:49–63
- He H, Garcia E (2009) Learning from imbalanced data. *IEEE Trans Knowl Data Eng* 21:1263–1284
- Kruhlik NL, Choi SS, Contrera JF, Weaver JL, Willard JM, Hastings KL, Sancilio LF (2008) Development of a phospholipidosis database and predictive quantitative structure-activity relationship (QSAR) models. *Toxicol Mech Methods* 18:217–227
- NCATS (2016) “Tox21 Data Browser.” <https://tripod.nih.gov/tox21>. Accessed 12 Feb 2017
- PubChem (2013) “Tox21 Phase II compound collection” <https://www.ncbi.nlm.nih.gov/pcsubstance/?term=tox21>. Accessed 12 Feb 2017
- Reasor MJ, Kacew S (2001) Drug-induced phospholipidosis: are there functional consequences? *Exp Biol Med* (Maywood) 226:825–830
- Shahane SA, Huang R, Gerhold D, Baxa U, Austin CP, Xia M (2014) Detection of phospholipidosis induction: a cell-based assay in high-throughput and high-content format. *J Biomol Screen* 19:66–76
- Slavov S, Pearce B, Buzatu D, Wilkes J, Beger R (2013) Complementary PLS and KNN algorithms for improved 3D-QSDAR consensus modeling of AhR binding. *J Cheminform* 5:47–62
- Slavov SH, Wilkes JG, Buzatu DA, Kruhlik NL, Willard JM, Hanig JP, Beger RD (2014) Computational identification of a phospholipidosis-toxicophore using ¹³C and ¹⁵N NMR-distance based fingerprints. *Bioorg Med Chem* 22:6706–6714
- Stoyanova-Slavova IB, Slavov SH, Buzatu DA, Beger RD, Wilkes JG (2017) 3D-SDAR modeling of hERG potassium channel affinity: a case study in model design and toxicophore identification. *J Mol Graph Model* 72:246–255
- Sun H, Xia M, Shahane SA, Jadhav A, Austin CP, Huang R (2013) Are hERG channel blockers also phospholipidosis inducers? *Bioorg Med Chem Lett* 23:4587–4590
- Titus SA, Beacham D, Shahane SA, Southall N, Xia M, Huang R, Hooten E, Zhao Y, Shou L, Austin CP, Zheng W (2009) A new homogeneous high-throughput screening assay for profiling compound activity on the human ether-a-go-go-related gene channel. *Anal Biochem* 394:30–38
- Villoutreix BO, Taboureau O (2015) Computational investigations of hERG channel blockers: new insights and current predictive models. *Adv Drug Deliv Rev* 86:72–82
- Wang Y, Huang R (2016) Correction of microplate data from high throughput screening. In: Zhu H (ed) *High-throughput screening assays in toxicology*. Humana Press, Springer Science + Business Media, New York, pp 123–134
- Wang Y, Jadhav A, Southall N, Huang R, Nguyen DT (2010) A grid algorithm for high throughput fitting of dose-response curve data. *Curr Chem Genomics* 4:57–66
- Witchel HJ (2011) Drug-induced hERG block and long QT syndrome. *Cardiovasc Ther* 29:251–259
- Xia M, Shahane SA, Huang R, Titus SA, Shum E, Zhao Y, Southall N, Zheng W, Witt KL, Tice RR, Austin CP (2011) Identification of quaternary ammonium compounds as potent inhibitors of hERG potassium channels. *Toxicol Appl Pharmacol* 252:250–258