Published in final edited form as:

Methods Mol Biol. 2018 ; 1800: 275–285. doi:10.1007/978-1-4939-7899-1_13.

Computational toxicology methods in chemical library design and high-throughput screening hit validation

Kirk E. Hevener*

Department of Pharmaceutical Sciences, University of Tennessee Health Science Center, 881 Madison Avenue, Suite 571, Memphis, TN 38163 (USA)

Abstract

The discovery of molecular toxicity in a clinical drug candidate can have a significant impact on both the cost and timeline of the drug discovery process. Early identification of potentially toxic compounds during screening library preparation or, alternatively, during the hit validation process, is critical to ensure that valuable time and resources are not spent pursuing compounds that may possess a high propensity for human toxicity. This report focuses on the application of computational molecular filters, applied either pre- or post-screening, to identify and remove known reactive and/or potentially toxic compounds from consideration in drug discovery campaigns.

Keywords

Molecular toxicity; computational filter; high-throughput screening; virtual screening; library design; drug discovery

1. Introduction

Toxicity is now one of the leading causes of compound failure in clinical drug development. A recent analysis of drug candidate attrition from several large pharmaceutical companies showed that safety and toxicity are now the greatest sources of failure [1]. It is well known that the physicochemical properties of drug candidates are associated with their toxicological outcomes [2–4], and decades of medicinal chemistry experience have resulted in the identification of specific functional groups and chemical motifs that are strongly associated with toxicological issues *in vivo* (toxicophores), often due to a higher propensity for chemical reactivity [5,6]. The application of this knowledge in the screening and removal of potentially toxic compounds from consideration early in the drug discovery process will be a critical factor in efforts to lower drug candidate attrition rates and mitigate the high costs associated with attrition late in the drug discovery process.

To date, a variety of techniques and methods have been developed to predict potential toxicity in clinical drug candidates early in the discovery process. Two general methods are often employed. The first involves the application of computationally developed algorithms

*To whom correspondence should be addressed. Phone: 901-448-1474. Fax: 901-448-6828. khevener@uthsc.edu.

or models to the identification and elimination of potentially toxic compounds from screening sets. These methods include quantitative structure-toxicity relationship models (QSTR), machine learning and pattern recognition techniques, toxicophore-mapping, and knowledge-based approaches [7–15]. The second method commonly employed is the application of chemical filters to screening sets designed to eliminate compounds from screening that have undesirable physicochemical properties (non-leadlikeness) and/or possess known or suspected toxicophores. A variety of filters have been developed, published, and successfully employed including early non-leadlike rules and properties [16], the Bristol-Myers Squibb published filters [17], the Eli Lilly rules for identifying potentially reactive compounds [6], the REOS filters (rapid elimination of swill) developed by Vertex [18], and the Astra Zeneca filters [19]. Additionally, there are several online structural alert web servers (e.g. ToxAlerts) that can be employed pre- or post-screening to flag potentially problematic compounds [20].

As a common underlying mechanism of *in vivo* toxicity is chemical reactivity, this report will primarily focus on the application of specific filters for prediction of chemically reactive, and thereby potentially toxic, functional groups or chemical motifs in the development of screening libraries, pre-filtering of libraries prior to virtual (HTVS) or high-throughput screening (HTS), and in post-HTS/HTVS hit confirmation studies. Chemical filters for leadlike properties (Rule of 5, etc.), drug-like characteristics, and those that influence desirable pharmacokinetic properties, will not be explicitly discussed herein [21–23]. Further, rules for predicting activity assay promiscuity (frequent hitters), e.g. PAINS (Pan-Assay Interference Compounds), are not explicitly discussed [24,25]. While there may be some overlap, a frequent hitting compound does not necessarily result in human toxicity, rather the hit frequency may be related to assay interference. Finally, it should be noted that chemical reactivity does not necessarily predict human toxicity, and many of the functional groups detailed below have not explicitly been linked to human toxicities, as such. However, their higher propensity for reactivity with biological macromolecules, i.e. protein acylation, lead toward a higher predictivity toward adverse human effects in general, as well as common assay interference, and it is recommended that they be removed from screening libraries if there is not sufficient reason to retain them, such as biochemical screens for covalently acting agents.

2. Applications

2.1. Screening library development tools

Consideration of potentially reactive or toxic compounds should be done as early as possible, including the screening library design stage. Many physicochemical properties are considered in the design of a chemical library. Properties influencing ‘drug-likeness’, favorable pharmacokinetic properties, molecular diversity (if a targeted library is not the goal), potential assay interference, synthetic accessibility, and chemical reactivity are commonly evaluated and used in the decision process to retain or remove compounds when building a small-molecule screening library [26]. Several tools are available to aid researchers in the identification of potentially toxic compounds while developing screening

libraries, including predictive tools and chemical filters based upon known or suspected reactive or toxic chemical moieties.

Among the research tools that can be directly applied to chemical library design, FAF-*Drugs4* (Free ADME-Tox Filtering Tool) is a useful online program that can be used to pre-screen chemical libraries during development or prior to high-throughput virtual or experimental screening [27]. Key features include the ability of the software to computationally predict ADME (Absorption, Distribution, Metabolism, & Excretion) properties and to remove salts, counterions, and duplicate compounds during the library design stage. The server includes a large set of pre-defined toxicophore filters and allows users to custom design their own filters as well.

The ZINC15 database presents a valuable resource to the drug discovery community for screening library development [28]. Provided by the Irwin and Shoichet laboratories at University of California, San Francisco, the ZINC15 database curates more than 120 million commercially available compounds from nearly 400 vendors or other catalogs for virtual screening. Compounds can be downloaded in a docking-ready, 3D format pre-filtered for drug-likeness and to remove compounds with potentially problematic chemical structures, including known toxicophores. Molecular properties are annotated for each compound, allowing researchers to download custom screening sets, such as fragment-like or lead-like sets. Additionally, compounds with known activity have been annotated with biological data allowing for researchers to design focused or targeted screening libraries. A robust web interface and search tool allows for rapid compound scaffold and similarity searching. A potential disadvantage is that the database is somewhat dated (2015), thus screening compounds downloaded may not be readily available for purchase if vendor stocks have been recently depleted.

There have also been a large variety of software reported that allow for *in silico* combinatorial library design, many of which incorporate tools for the prediction of physicochemical properties and toxicity risk assessment. Representative examples include proprietary software such as Schrödinger's CombiGlide, QikProp, and LigPrep tools; Biovia's (Accelrys) Discovery Studio; the Design Module of MedChem StudioTM; and the open-source ChemT package from BioChemCore [29]. Table 1 provides a list of software and web-based tools for prediction of chemical reactivity and toxicity, many of which are applicable to the design of chemical screening libraries.

2.2 Tools for prediction of reactivity/toxicity

A wide variety of tools have been developed for the prediction of molecular toxicity, including predictive software, servers, and databases. A representative selection is discussed here, with a more complete listing, including system classification and web address given in Table 1. One example is the eTOX project. Sponsored by the European Innovative Medicines Initiative (IMI), the eTOX project collected toxicological data from pharmaceutical industry and academia into an online database that has been used to create a series of toxicity prediction models, including the online toxicity prediction system, eTOXsys [30]. Key features of this system are the ability to query the database for toxicities

related to the target rather than the drug and chemical substructures showing a high correlation with specific toxicities.

Several companies offer model-based toxicity prediction services, including Leadscape's® Toxicity Database comprised of over 180,000 compounds and over 400,000 toxicity study results. Leadscape® models include hepatobiliary, cardiological and urinary effects as well as developmental, genetic, and neurotoxicity. Derek Nexus (Lhasa Limited) is another member-based (proprietary) service that offers a rapid toxicity assessment for compounds submitted.

ToxAlerts is an open-source, web-based server integrated with the Online Chemical Modeling Environment (OCHEM) that collects and stores toxicological data collected from existing literature or submitted by users [20]. Structural alerts in the form of SMARTS patterns are generated and can be used to screen individual compounds for toxicity prediction or to pre-filter libraries during screening library design. Structure alerts currently exist for endpoints including mutagenicity, carcinogenicity, skin sensitization, and compounds that can form reactive metabolites.

Lastly, the U.S. National Library of Medicine's Toxicology Data Network (TOXNET) is a publicly available resource that allows users to screen specific compounds against a large variety of toxicology databases (HSDB, CCRIS, GENETOX, etc.) and literature references (TOXLINE, DART) [31,32]. More of a data mining than predictive tool, the TOXNET databases are useful for assessing hit compounds from virtual or experimental screening during the hit validation process to assess a compound for further advancement. A note of caution is merited here as the absence of data found for a submitted compound does not necessarily preclude a propensity for toxicity, rather the absence of a literature or database report.

2.3 Structure alerts for reactive and/or toxic functional groups

The use of structure alerts for the identification of potentially reactive or toxic compounds is regularly employed in both pre-filtering of chemical libraries during the library design stage and post-filtering of screening compounds during the hit validation stage [33–36]. There are advantages and disadvantages to both pre- and post-filtering strategies. For the former, the unilateral removal of compounds containing known reactive or toxic functional groups may result on the loss of a valid hit compound which might be modified during the optimization process to remove the offending chemical moiety. For the latter, there is an increased personnel and infrastructure expense related to the cost of screening larger libraries that have not been pre-filtered, as well as a potential high cost of late-stage failure of a clinical candidate. One possible solution is the use of customizable threshold cutoffs, an available option for most structure alert algorithms, based on the number of occurrences of a reactive substructure (e.g. no more than two nitro groups). Alternatively, toxicophore filtering may be cautiously employed post-screening to 'flag' compounds identified with potentially problematic groups that may warrant close attention. Lastly, it is possible to employ a pre-screen filter for the elimination of particularly high-risk compounds, coupled with a post-screen filter to flag lower risk compounds, such as PAINS compounds or compounds with

less reactive functional groups occasionally still seen in approved drugs (e.g. aniline & nitro groups).

In most cases, structure alert algorithms employ the use of SMARTS (SMILES Arbitrary Target Specification) patterns to identify pre-determined chemical patterns in compounds and chemical libraries. SMARTS, developed by Daylight Chemical Information Systems, Inc., is a SMILES-based 2D line notation that allows for the incorporation of variability, wildcards, atomic properties and connectivity in the search [37]. Using SMARTS, atoms can be represented by atomic number, capital or lower-case letters. As an example, carbon can be represented as **C** (aliphatic carbon atom), **c** (aromatic carbon atom), or **[#6]** (any carbon atom). Wildcard values can be included in SMARTS patterns to represent ***** (any atom), **a** (aromatic atom), **A** (aliphatic atom), **R** (ring membership), **r** (ring size), **X** (connectivity), charge, chirality, valence, mass, and several others. Values for atoms can be coupled together for greater specificity using brackets and semi-colons, as with **[C;X2]** (aliphatic carbon with two total bonds, including implicit hydrogens). Variability at atomic positions can be specified using brackets and commas, as with **[O,N,S;X2;r3]** (oxygen, nitrogen, or sulfur with two total bonds in a 3-membered ring system). Various symbols are used to represent atomic bonds making connections between atoms, including **-** (single bond), **=** (double bond), **#** (triple bond), **:** (aromatic bond), **~** (any bond), **@** (any ring bond), and others. A missing bond symbol is interpreted as 'single or aromatic', which can be used to prevent the possibility of missing an aromatic system described using double bonds. Positional branching is represented, as with SMILES, by parentheses and logical operators, **!** (not) and **&** (and) can be included for additional specificity. Recursive expressions and component-level grouping can also be incorporated into SMARTS patterns. Interested readers are referred to the DAYLIGHT website for additional details and manuals [37].

3. Summary of Reactive Structure Filters

As mentioned above, several groups have published useful rules and filters that have been successfully employed in the library design and screening processes, and there have been several helpful reviews published in the area. The seventh edition of Burger's Medicinal Chemistry, Drug Discovery, and Development includes a very useful chapter summarizing structural alerts for toxicity [34]. Structure alerts are categorized by compounds with inherent chemical reactivity (e.g., acylating & alkylating groups), compounds requiring metabolism to generate a reactive compound (e.g., anilines, nitro-aromatics, hydrazines), compounds exhibiting CYP450 interference (e.g., imidazoles, triazoles, 2,6-unsubstituted pyridines), and compounds exhibiting a high DNA binding propensity (e.g., polycyclic aromatic compounds).

Several manuscripts have been published by Rishton discussing non-leadlikeness and compound reactivity in drug discovery and include several suggested functional groups that may lead to false positives or potential toxicity [5,16]. Reactive groups explicitly detailed include sulfonyl, acyl and alkyl halides, anhydrides, aldehydes, imines, epoxides, sulfonate and phosphonate esters, Michael acceptors, and several others. Pearce and coworkers at Bristol-Myers Squibb published a useful set of filters for use in the design of high-throughput screening libraries [17]. The filters designed by this group were categorized as

exclusion filters, flagging compounds for removal from the library, and information filters, annotating potentially problematic compounds but not removing them. Bruns and coworkers at Eli Lilly recently published a similar set of filters and described a demerit-based system employed at Lilly to identify potentially problematic compounds in their screening sets. They describe 275 rules for the identification of reactive compounds, compounds that may interfere with assays (e.g., fluorescence, absorbance, quenching), compounds with intrinsic protein damaging capability (e.g., oxidizing agents and detergents), unstable compounds, and compounds lacking drug-like features. The REOS (Rapid Elimination of Swill) rules described by Walters and coworkers at Vertex Pharmaceuticals includes a set of more than 200 functional group filters, which include reactive and other undesirable functional groups [18,38]. Lastly, researchers at Astra Zeneca have recently published a set of filters (AZ-Filters) for use in library design and screening hit validation [19]. The AZ filters include patterns for “bland structures” (primarily non-druglike features), reactive structures (potentially toxicity), frequent hitters, dyes, natural products, and others.

Table 2 includes a listing of commonly filtered functional groups associated with reactivity or toxicity and their associated SMARTS patterns. The table represents functional groups common to all the publications discussed above and is not intended to present an all-encompassing list. Readers are referred to the resources discussed above for additional structural alerts and filters that may be applicable to their specific research. Representative chemically reactive groups commonly included in structure alerts or filters include acylating and alkylating agents, aldehydes and ketones, Michael acceptors, reactive esters and thioesters, anhydrides, imines, cyanates and cyanides. Other known or suspected toxicophores that are often included in filters are quinones (or quinone-forming groups), nitroaromatic and aromatic amines, nitrosamines, acylhydrazides, thioureas, sulfur & nitrogen mustards, polycyclic aromatic systems, triazenes, epoxides & aziridines, and aminopyrines [39,40]. Additionally, many structure alert filters are employed to remove groups associated both with non-drug-likeness as well as potential toxicity, including the lanthanides, actinides, noble gases, alkali metals, alkali earth metals, and transition metals. Lastly, common protecting groups, chemical reagents and intermediates (triflate, esters of hydroxybenzotriazole, Lawesson’s reagent, pentafluorophenyl esters, chloramidine, triacyloxime, reactive cyanides, reactive azo groups, etc.) should be considered for removal by filters prior to screening or in the library design stage. It should be noted that there is considerable overlap between the reactive and toxicophoric moieties discussed here and groups associated with promiscuity, assay interference, and general non-leadlikeness.

4. Conclusions

The early identification of potential toxicity in drug discovery campaigns is critical to prevent significant financial loss resulting from pursuing compounds that may fail in late stage development. Methods introduced here included the application of resources such as servers, databases, and predictive algorithms to identify or flag potentially toxic compounds based upon their physicochemical properties and the use of structure alerts, or functional group filters, to identify compounds based upon the presence of known reactive functional groups or toxicophores. Both strategies can be employed in the library design stage as well as the hit validation stage and can be used pre- or post-screening, depending on the goals of

the program and the resources available. Finally, the strategies discussed here are applicable both to virtual and experimental library screening.

Acknowledgments

This work was supported by NIH grant AI126755 and faculty development program funding from UTHSC College of Pharmacy to K.E.H.

References

1. Waring MJ, Arrowsmith J, Leach AR, Leeson PD, Mandrell S, Owen RM, Pairaudeau G, Pennie WD, Pickett SD, Wang J, Wallace O, Weir A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discovery*. 2015; 14(7):475–486. DOI: 10.1038/nrd4609 [PubMed: 26091267]
2. Hughes JD, Blagg J, Price DA, Bailey S, Decrescenzo GA, Devraj RV, Ellsworth E, Fobian YM, Gibbs ME, Gilles RW, Greene N, Huang E, Krieger-Burke T, Loesel J, Wager T, Whiteley L, Zhang Y. Physicochemical drug properties associated with in vivo toxicological outcomes. *Bioorg Med Chem Lett*. 2008; 18(17):4872–4875. DOI: 10.1016/j.bmcl.2008.07.071 [PubMed: 18691886]
3. Price DA, Blagg J, Jones L, Greene N, Wager T. Physicochemical drug properties associated with in vivo toxicological outcomes: a review. *Expert Opin Drug Metab Toxicol*. 2009; 5(8):921–931. DOI: 10.1517/17425250903042318 [PubMed: 19519283]
4. Barratt MD. Prediction of toxicity from chemical structure. *Cell Biol Toxicol*. 2000; 16(1):1–13. [PubMed: 10890502]
5. Rishton GM. Reactive compounds and in vitro false positives in HTS. *Drug Discovery Today*. 1997; 2(9):382–384.
6. Bruns RF, Watson IA. Rules for identifying potentially reactive or promiscuous compounds. *Journal of medicinal chemistry*. 2012; 55(22):9763–9772. DOI: 10.1021/jm301008n [PubMed: 23061697]
7. Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. *J Comput Chem*. 2017; 38(16):1291–1307. DOI: 10.1002/jcc.24764 [PubMed: 28272810]
8. Gawehn E, Hiss JA, Schneider G. Deep Learning in Drug Discovery. *Mol Inform*. 2016; 35(1):3–14. DOI: 10.1002/minf.201501008 [PubMed: 27491648]
9. van de Waterbeemd H, Gifford E. ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discovery*. 2003; 2(3):192–204. DOI: 10.1038/nrd1032 [PubMed: 12612645]
10. Bugrim A, Nikolskaya T, Nikolsky Y. Early prediction of drug metabolism and toxicity: systems biology approach and modeling. *Drug Discovery Today*. 2004; 9(3):127–135. DOI: 10.1016/S1359-6446(03)02971-4 [PubMed: 14960390]
11. Segall MD, Barber C. Addressing toxicity risk when designing and selecting compounds in early drug discovery. *Drug Discovery Today*. 2014; 19(5):688–693. DOI: 10.1016/j.drudis.2014.01.006 [PubMed: 24451294]
12. Gertrudes JC, Maltarollo VG, Silva RA, Oliveira PR, Honorio KM, da Silva AB. Machine learning techniques and drug design. *Curr Med Chem*. 2012; 19(25):4289–4297. [PubMed: 22830342]
13. Moroy G, Martiny VY, Vayer P, Villoutreix BO, Miteva MA. Toward in silico structure-based ADMET prediction in drug discovery. *Drug Discovery Today*. 2012; 17(1–2):44–55. DOI: 10.1016/j.drudis.2011.10.023 [PubMed: 22056716]
14. Gini G. QSAR Methods. *Methods Mol Biol*. 2016; 1425:1–20. DOI: 10.1007/978-1-4939-3609-0_1 [PubMed: 27311459]
15. Singh PK, Negi A, Gupta PK, Chauhan M, Kumar R. Toxicophore exploration as a screening technology for drug design and discovery: techniques, scope and limitations. *Arch Toxicol*. 2016; 90(8):1785–1802. DOI: 10.1007/s00204-015-1587-5 [PubMed: 26341667]
16. Rishton GM. Nonleadlikeness and leadlikeness in biochemical screening. *Drug Discovery Today*. 2003; 8(2):86–96. [PubMed: 12565011]

17. Pearce BC, Sofia MJ, Good AC, Drexler DM, Stock DA. An empirical process for the design of high-throughput screening deck filters. *J Chem Inf Model*. 2006; 46(3):1060–1068. DOI: 10.1021/ci050504m [PubMed: 16711725]
18. Walters WP, Ajay, Murcko MA. Recognizing molecules with drug-like properties. *Curr Opin Chem Biol*. 1999; 3(4):384–387. [PubMed: 10419858]
19. Cumming JG, Davis AM, Muresan S, Haerberlein M, Chen H. Chemical predictive modelling to improve compound quality. *Nat Rev Drug Discovery*. 2013; 12(12):948–962. DOI: 10.1038/nrd4128 [PubMed: 24287782]
20. Sushko I, Salmina E, Potemkin VA, Poda G, Tetko IV. ToxAlerts: a Web server of structural alerts for toxic chemicals and compounds with potential adverse reactions. *J Chem Inf Model*. 2012; 52(8):2310–2316. DOI: 10.1021/ci300245q [PubMed: 22876798]
21. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods*. 2000; 44(1):235–249. S1056-8719(00)00107-6 [pii]. [PubMed: 11274893]
22. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. *Journal of medicinal chemistry*. 2002; 45(12):2615–2623. jm020017n [pii]. [PubMed: 12036371]
23. Teague SJ, Davis AM, Leeson PD, Oprea T. The Design of Leadlike Combinatorial Libraries. *Angew Chem Int Ed Engl*. 1999; 38(24):3743–3748. [pii]. DOI: 10.1002/(SICI)1521-3773(19991216)38:24<3743::AID-ANIE3743>3.0.CO;2-U [PubMed: 10649345]
24. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *Journal of medicinal chemistry*. 2010; 53(7):2719–2740. DOI: 10.1021/jm901137j [PubMed: 20131845]
25. Dahlin JL, Nissink JW, Strasser JM, Francis S, Higgins L, Zhou H, Zhang Z, Walters MA. PAINS in the assay: chemical mechanisms of assay interference and promiscuous enzymatic inhibition observed during a sulfhydryl-scavenging HTS. *Journal of medicinal chemistry*. 2015; 58(5):2091–2113. DOI: 10.1021/jm5019093 [PubMed: 25634295]
26. Dolle RE. Historical overview of chemical library design. *Methods Mol Biol*. 2011; 685:3–25. DOI: 10.1007/978-1-60761-931-4_1 [PubMed: 20981516]
27. Lagorce D, Sperandio O, Baell JB, Miteva MA, Villoutreix BO. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res*. 2015; 43(W1):W200–207. DOI: 10.1093/nar/gkv353 [PubMed: 25883137]
28. Sterling T, Irwin JJ. ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model*. 2015; 55(11):2324–2337. DOI: 10.1021/acs.jcim.5b00559 [PubMed: 26479676]
29. Abreu RM, Froufe HJ, Daniel PO, Queiroz MJ, Ferreira IC. ChemT, an open-source software for building template-based chemical libraries. *SAR QSAR Environ Res*. 2011; 22(5–6):603–610. DOI: 10.1080/1062936X.2011.604097 [PubMed: 21846264]
30. Sanz F, Carrio P, Lopez O, Capoferri L, Kooi DP, Vermeulen NP, Geerke DP, Montanari F, Ecker GF, Schwab CH, Kleinoder T, Magdziarz T, Pastor M. Integrative Modeling Strategies for Predicting Drug Toxicities at the eTOX Project. *Mol Inform*. 2015; 34(6–7):477–484. DOI: 10.1002/minf.201400193 [PubMed: 27490391]
31. Fowler S, Schnall JG. TOXNET: information on toxicology and environmental health. *Am J Nurs*. 2014; 114(2):61–63. DOI: 10.1097/01.NAJ.0000443783.75162.79
32. Wexler P. TOXNET: an evolving web resource for toxicology and environmental health information. *Toxicology*. 2001; 157(1–2):3–10. [PubMed: 11164971]
33. Zhu T, Cao S, Su PC, Patel R, Shah D, Chokshi HB, Szukala R, Johnson ME, Hevener KE. Hit identification and optimization in virtual screening: practical recommendations based on a critical literature analysis. *Journal of medicinal chemistry*. 2013; 56(17):6560–6572. DOI: 10.1021/jm301916b [PubMed: 23688234]
34. Blagg J. Structural Alerts for Toxicity. In: Abraham DJ, Rotella DP, editors *Burger's Medicinal Chemistry and Drug Discovery*. 7. John Wiley & Sons, Inc; 2010. 301–334.
35. Smith GF. Designing drugs to avoid toxicity. *Prog Med Chem*. 2011; 50:1–47. DOI: 10.1016/B978-0-12-381290-2.00001-X [PubMed: 21315927]

36. Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry*. 2005; 48(1):312–320. DOI: 10.1021/jm040835a [PubMed: 15634026]
37. SMARTS - A Language for Describing Molecular Patterns. Daylight Chemical Information Systems, Inc; <http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html> [Accessed December 20 2017]
38. Walters WP, Stahl MT, Murcko MA. Virtual screening - an overview. *Drug Discovery Today*. 1998; 3(4):160–178.
39. Williams DP, Naisbitt DJ. Toxicophores: groups and metabolic routes associated with increased safety risk. *Current opinion in drug discovery & development*. 2002; 5(1):104–115. [PubMed: 11865664]
40. Hakimelahi GH, Khodarahmi GA. The Identification of Toxicophores for the Prediction of Mutagenicity, Hepatotoxicity and Cardiotoxicity. *Journal of the Iranian Chemical Society*. 2005; 2(4):244–267.

Table 1
Selected Software and Web-Based Tools for Prediction of Chemical Reactivity/Toxicity

Software/Tool	Classification	Website or Web Reference
ACD/Percepta	Software	http://www.acdlabs.com/products/percepta/predictors.php
ADMET Predictor™	Software	http://www.simulations-plus.com/software/admet-property-prediction-qsar/
BIOVIA Discovery Studio	Software	http://accelrys.com/products/collaborative-science/biovia-discovery-studio/
CompuDrug/HazardExpertPro	Software	http://www.compuDrug.com/?q=node/35
Derek Nexus	Software	https://www.hasalimited.org/products/derek-nexus.htm
Leadscope®	Software	http://www.leadscope.com/model_appliers/
MultiCASE/CASE Ultra	Software	http://www.multicase.com/case-ultra
Schrödinger/QuickProp	Software	https://www.schrodinger.com/qikprop
Way2Drug GUSAR	Software	http://www.way2drug.com/mg/about.php
FAF-Drugs4	Server	http://mobyli.e.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py?form=admetox#forms::FAF-Drugs4
eTOXsys	Server	https://etoxsys.eu/etoxsys.v3-demo-bk/dashboard/
Mecule/Toxicity Checker	Server	https://mecule.com/apps/toxicity-checker/
PASS Online	Server	http://pharmaexpert.ru/Passonline/index.php
OpenTox/ToxPredict	Server	http://www.opentox.net/library/toxicity-prediction
ToxAlert	Server	https://ochem.eu/home/show.do
ToxiPred	Server	http://crdd.osdd.net/oscadd/toxipred/
VirtualToxLab™	Server	http://www.biograf.ch/index.php?id=projects&subid=virtualtoxlab
Lazar	Server	https://lazar.in-silico.de/predict
Leadscope®	Database	http://www.leadscope.com/product_info.php?products_id=78
NIH/NLM TOXNET	Database	https://toxnet.nlm.nih.gov/
SuperToxic	Database	http://bioinformatics.charite.de/supertoxic/index.php?site=home

Table 2

Commonly Filtered Reactive Groups and their SMARTS Patterns

Group Name	SMARTS Patterns ^I
1,2-Dicarbonyls	[C;X3](=O)([C;X3](=O))
Acyl Halides	[F,Cl,Br,I][C;X3]=[O,S]
Aldehydes	[#6][C;H1]=[O;X1]
Alkyl Halides, P/S Halides, Mustards	[Cl,Br,I][P,S,C&H2&X4,C&H3&X4]
Alkyl Sulfonates, Sulfate Esters	[#6]O[S;X4](=O)=O
Alpha-halocarbonyls	[#6][C;X3](=[O;X1])-[C;H1,H2]-[F,Cl,Br,I]
Alpha-Beta Unsaturated Nitriles	[#6]=[#6]C#N
Anhydride	[O;X2]([CX3,S,P]=O)([CX3,S,P]=O)
Azides	[#6][N;X2]=[N;X2]=[N;X1]
Beta-Carbonyl Quaternary Nitrogen	[C;X3](=O)[C][N,n;X4]
Beta-Heterosubstituted Carbonyls	[O;X1]=C[C;H2]C[F,Cl,Br,I]
Carbodiimides	[#6][N;X2]=[C;X2]=[N;X2][#6]
Diazos, Diazoniums	[#6]~[NX2;+0,+1]~[NX1;-1,+1,+0]
Disulfides	[S;X2]~[S;X2]
Epoxides, Thioepoxides, Aziridines	[O,N,S;X2;r3](C)C
Formates	[O;X2][C;H1]=O
Halopyrimidines	[F,Cl,Br,I]c(nc)nc
Heteroatom-Heteroatom Single Bonds	[O,N,S;X2]~[O,N,S;X2]
Imines (Schiff's Base)	[N;X2]([!#1])=[C;X3][C;H2,H3]
Isocyanates, Isothiocyanates	[#6][N;X2]=C=[O,S&X1]
Isonitriles	[#6][N;X2]#[C;X1]
Michael Acceptors	[#6]=[#6][#6,#16]=[O]
Nitroaromatic	[c;X3][\$([NX3](=O)=O),\$([NX3+](=O)[O-])]
Nitrosos, Nitrosamines	[#6,#7][N;X2](=O)
Perhalomethylketones	[#6][C;X3](=O)[C;X4]([F,Cl,Br,I])([F,Cl,Br,I])[F,Cl,Br,I]
Phosphines, Phosphoranes	[#6][#15&X3,#15&X5]([#6])~[#6]
Phosphinyl halides	[P;X3][Cl,Br,I]
Reactive cyanides	N#C[C&X4,C&X3]~[O&X1,O&H1&X2]
Sulfenyl, Sulfinyl, Sulfonyl halides	[F,Cl,Br,I][\$([SX2]),\$([S;X3]=O),\$([S;X4](=O)=O)]
Thiocyanate	[#6][S]C#N
Thioesters	*SC(=O)*
Thiourea	[SX1]~C([N&!R&X3,N&!R&X2])[N&!R&X3,N&!R&X2]
Vinyl halides	[C;X2]=[C;X2]-[F,Cl,Br,I]

^I SMARTS patterns were generated using the Schrödinger Ligand Filtering utility and verified using SMARTSviewer (<http://smartsview.zbh.uni-hamburg.de/>).