# ToxAlerts: A Web Server of Structural Alerts for Toxic Chemicals and Compounds with Potential Adverse Reactions

Iurii Sushko,[†] Elena Salmina,[‡] Vladimir A. Potemkin,[§] Gennadiy Poda,[⊥] and Igor V. Tetko*,[†,||]

[†]eADMET GmbH, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

[‡]Institute of Organic Synthesis of the Ural Department of the Russian Academy of Sciences, 620069 Ekaterinburg, Russia

[§]Pharmaceutical Chemistry, Chelyabinsk State Medical Academy, Chelyabinsk, Russian Federation 454048

[⊥]Medicinal Chemistry Platform, Ontario Institute for Cancer Research, 101 College Street, MaRS Centre, South Tower, Toronto, Ontario, Canada M5G 0A3

[||]Institute of Structural Biology, Helmholtz Zentrum Muenchen, German Research Center for Environmental Health, Ingolstädter Landstraße 1, b. 60w D-85764 Neuherberg, Germany

**ABSTRACT:** The article presents a Web-based platform for collecting and storing toxicological structural alerts from literature and for virtual screening of chemical libraries to flag potentially toxic chemicals and compounds that can cause adverse side effects. An alert is uniquely identified by a SMARTS template, a toxicological endpoint, and a publication where the alert was described. Additionally, the system allows storing complementary information such as name, comments, and mechanism of action, as well as other data. Most importantly, the platform can be easily used for fast virtual screening of large chemical datasets, focused libraries, or newly designed compounds against the toxicological alerts, providing a detailed profile of the chemicals grouped by structural alerts and endpoints. Such a facility can be used for decision making regarding whether a compound should be tested experimentally, validated with available QSAR models, or eliminated from consideration altogether. The alert-based screening can also be helpful for an easier interpretation of more complex QSAR models. The system is publicly accessible and tightly integrated with the Online Chemical Modeling Environment (OCHEM, http://ochem.eu). The system is open and expandable: any registered OCHEM user can introduce new alerts, browse, edit alerts introduced by other users, and virtually screen his/her data sets against all or selected alerts. The user sets being passed through the structural alerts can be used at OCHEM for other typical tasks: exporting in a wide variety of formats, development of QSAR models, additional filtering by other criteria, etc. The database already contains almost 600 structural alerts for such endpoints as mutagenicity, carcinogenicity, skin sensitization, compounds that undergo metabolic activation, and compounds that form reactive metabolites and, thus, can cause adverse reactions. The ToxAlerts platform is accessible on the Web at http://ochem.eu/alerts, and it is constantly growing.
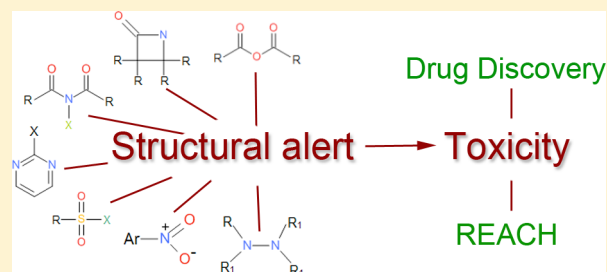
## INTRODUCTION

The identification of potentially toxic chemical compounds represents an important problem. This problem is crucial, for example, in early stages of drug discovery. Withdrawing a launched drug from the market due to adverse drug reactions (ADRs) is the costliest failure of a drug discovery program, frequently followed by years and years of lawsuits and costly settlements. Thus, it is paramount for medicinal chemists to be aware of potential structural features that may cause ADRs at efficacious doses. Another related field is the environmental toxicology (or ecotoxicology), which assesses the environmental risk of industrially produced chemicals. Again, the studies to perform such estimation are based on expensive and time-consuming experiments.

One approach to the toxicity assessment is using predictions provided by the quantitative structure−activity relationship (QSAR) models, which use different machine learning methods to correlate structural properties of chemicals (or so-called

molecular descriptors) to their toxicity.[1] Although QSAR modeling is a powerful technique, the two major problems are the applicability domain and interpretability of such models. The models based on powerful methods, such as artificial neural networks (ANN) or support vector machines (SVM), frequently use hundreds of descriptors and are particularly difficult to interpret. The lack of interpretability makes it difficult for these models to be used in the design of new potential drugs and to comply with the rules (so-called OECD rules) during the validation step.[2]

Another approach that is a lot easier to interpret is related to molecular patterns that are associated with particular types of toxicity or ADRs either directly or after undergoing of a metabolic activation *in vivo* (or bioactivation). These structural features are known as "toxicophores" or "structural alerts"
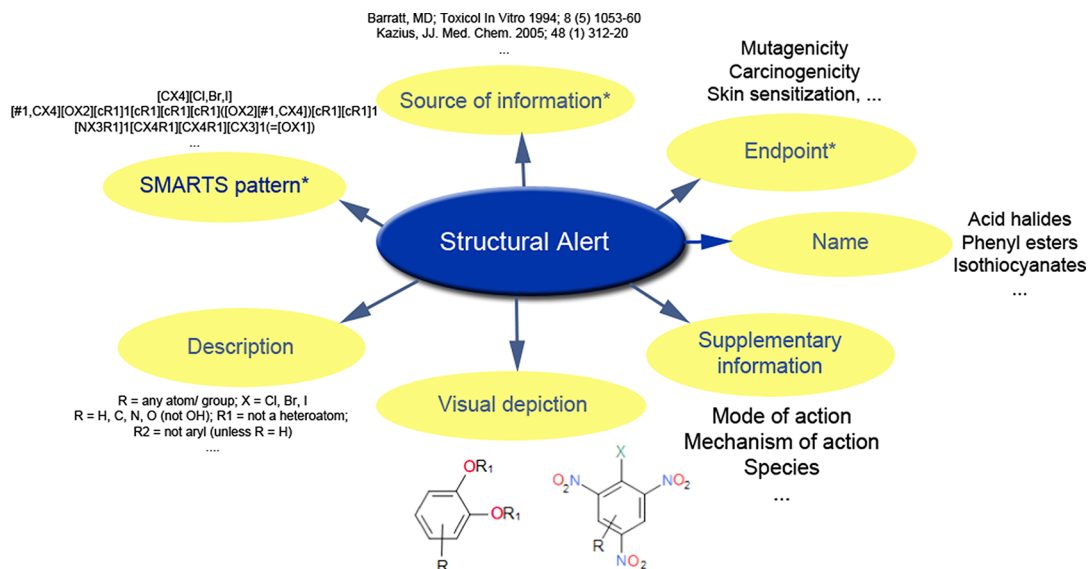
**Figure 1.** Simplified schema of the ToxAlerts database. The key identifying elements are marked with a star. In addition, the alerts have several other properties, i.e. the name of the OCHEM user who introduced or modified the data, creation date and time, modification history, and visibility control inherited from OCHEM.

(SAs) and represent a very simple and easily interpretable way to flag compounds with potential toxicity. Well-known expert systems based on SAs are the MultiCASE Expert Systems from MultiCASE, Inc.[3] and DEREK from LHASA.[4]

The studies performed in the past decade have shown that structural alerts is an efficient technique to detect potentially toxic chemicals.[5−16] Screening chemical compounds against known structural alerts can be a good practice to complement the QSAR models and to help interpret their predictions.

To make the wealth of knowledge of the structural alerts that could cause ADRs and toxic chemicals, we have developed ToxAlerts (http://ochem.eu/alerts), an open Web-based platform for uploading and storing structural alerts published in scientific literature in a structured manner with a capability to virtually screen compound libraries against these alerts to flag toxic chemicals and compounds with potential ADRs. The major goal of the project was to develop an open, community-driven, and expandable system.

## PLATFORM DESCRIPTION

ToxAlerts platform consists of two main components: a database to store the structural alerts for various toxicological endpoints in an organized manner and a facility to virtually screen chemical structures against these alerts.

**Database Structure.** The central entity in the ToxAlerts database is a structural alert (a substructure or, more universally, a SMARTS string), which is uniquely identified by the following:

1. A structural pattern represented by a SMiles ARbitrary Target Specification (SMARTS) string.
2. A publication, where the alert was mentioned.
3. A toxicological endpoint associated with this alert (e.g., carcinogenicity or skin sensitization).

In addition to these key components, the database is designed to store any other additional information, such as:

1. A chemical name of the compound class represented by the alert (e.g., "Acid halides", "Sulphonyl azides", etc.).

2. A visual depiction of the alert: since an automatic generation of a depiction from a pattern is an ambiguous and a nontrivial task, users can generate and upload their depictions in PNG format manually.
3. Position of the alert in the publication (i.e., page, table, line).
4. Arbitrary supplementary information (e.g., mechanism of action associated with the alert, species, metabolic activation information, etc.).

A simplified schema of the database is presented in Figure 1.

*Patterns of Structural Alerts using SMARTS Strings Representation.* As it was mentioned above, ToxAlerts uses SMARTS patterns to represent toxicological and ADRs structural alerts. The major advantage of using SMARTS is that a compound can be matched against the alert in an automatic manner using one of the available chemical libraries. For our purposes, we use the MolSearch utility from ChemAxon.[17] Although SMARTS are not always easily interpretable, the visual depiction, the chemical name, and the description provide a nice way to understand the alert.

*Extended Pattern Syntax.* In some cases, SMARTS patterns may not be sufficient to represent a desired alert. ToxAlerts provides an extended syntax, which supports logical operations (conjunction and inversion), restrictions on molecular weight, and on count of molecular fragments. In brief, a user can write expression like

$$MW > 100 \text{ AND}$$
$$[\$([NX3](=O)=O), \$([NX3+](=O)[O−])][!\#8] > 2$$

to match molecules with molecular weight of more than 100 Da and containing more than two nitro groups. Additionally, ToxAlerts supports variable substitutions (also known as "vector bindings", http://www.daylight.com/dayhtml/doc/prog/prog.smarts.html#9.3), which allow to simplify complex patterns and make them more interpretable. The list of available variables is managed centrally by the ToxAlerts moderators and can be extended on users' request.
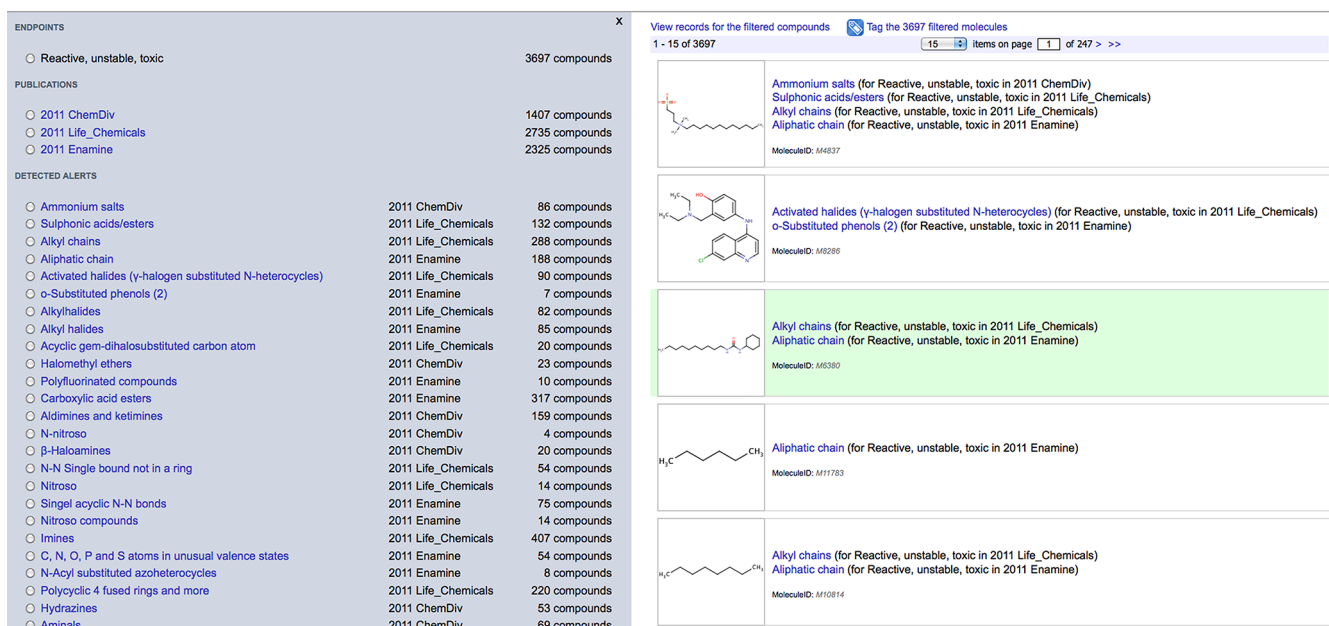
**Figure 2.** ToxAlerts dialogue with virtual screening results. The left panel shows the compounds grouped by alerts and endpoints. The right panel shows the identified compounds and the respective alerts.

*Source of Information.* It is a strict policy of the ToxAlerts database to store the structural alerts together with the source of information, e.g. a scientific publication or a book reference. This policy makes it easier to moderate and curate the database and avoid duplications. This way the end-users have the complete information about the source and origin of the alerts and, thus, can better decide how to proceed in each particular case. It is also possible to specify a "dummy" unpublished article for yet to be published data.

*Data Upload.* Structural alerts can be uploaded from an Excel file: one alert per row and one alert component per column. The structure of the Excel file is simple and described in detail on the data upload page: http://ochem.eu/alerts/upload.

A very convenient feature is a possibility to select structural alerts using particular criteria (e.g., endpoint type, publication or alert name) and to group them into sets that can be saved individually by each user under a given name. Such sets of alerts can be used for virtual screening of libraries to flag toxic and compounds that could cause ADRs.

**Integration with Online Chemical Modeling Environment.** The Online Chemical Modeling Environment (OCHEM) is a platform for storing experimentally measured properties and activities of chemical compounds and for development of QSAR models.[18] OCHEM is a collaborative, user-friendly resource that could be shared by users: any user on the Web can register, introduce new data, and develop QSAR models as well as assess published data and models introduced by other users. The database is tightly integrated into the QSAR modeling framework. The experimental data stored in the database can be easily manipulated to create sets to build predictable QSAR models using a variety of machine learning techniques (e.g., neural networks, support vector machines, multivariate linear regressions, etc.).

ToxAlerts is tightly integrated into OCHEM: the platforms share the same database of users, publications, and chemical structures. Moreover, ToxAlerts promotes the same philoso-phy: the data uploaded by a given user and can be freely accessed by other users. In addition, both ToxAlerts and OCHEM strictly require all data to be complemented with the original source of information—a reference to a scientific publication.

Integration with OCHEM gives many benefits. First, the compound sets (referred to as "baskets" and "tags") prepared in the OCHEM database can be readily used for screening against structural alerts in ToxAlerts. Second, the compounds filtered by alerts can be exported in a wide variety of formats and for any further calculation and analysis. Third, the presence or absence of particular structural alerts can be used as molecular descriptors for the development of QSAR models to predict ADRs.

Furthermore, ToxAlerts uses the system for distributed calculations facilitated by OCHEM. This not only allows to increase the speed of screening by 20−30 times but also to postpone fetching the screening results: all the running calculations can be managed and tracked *via* a centralized interface accessible for a logged-in user.

**Virtual Screening of Compound Libraries.** ToxAlerts functionality is not only limited to storing structural alerts. The alerts available in the database can be used for virtual screening of compound libraries to flag potentially toxic compounds. The results of screening can be saved by users and/or uploaded in the OCHEM database in SDF and SMILES formats.

ToxAlerts provides several ways of the structural alerts selection: screening against all available alerts, a possibility to select alerts for a particular endpoint or alerts from a particular publication, or use a custom previously saved set of alerts.

The screening procedure detects all the compounds that match at least one selected alert. The summary is reported with compounds grouped by alerts and endpoints. Furthermore, the results can be also filtered by a specific endpoint or alert.

An example dialogue that displays the screening results is presented in Figure 2.

The identified potentially toxic compounds can be exported in a wide variety of formats or analyzed further in OCHEM platform. Exemplary use cases include the following:

1. Mark all the identified compounds with a particular OCHEM tag and use this tag to predict the compounds with available QSAR models.
2. Find experimental measurements for the selected compounds, which can be further used to estimate specificity of each toxicological alert.

Thus the screening ToxAlerts facility is a powerful utility for identification of potentially toxic chemicals and for processing the results in a flexible manner due to a tight integration with the OCHEM platform.

**Data Quality Standards.** ToxAlerts is a community-driven resource and virtually anyone on the Web can introduce and modify the structural alerts in the database. Therefore, there are issues regarding data quality and safety. ToxAlerts addresses these issues in several ways.

*Validated Users.* All ToxAlerts users are categorized into several types: guests, registered users, validated users, and moderators and their privileges increase correspondingly. The user categories are assigned manually by the OCHEM administration team. To enhance data safety, data introduced by validated users can be modified only by validated users or moderators. Moreover, ToxAlerts logs the users who created or modified the data as well as the history of all the changes. The concept of validated users partially addresses the data quality issue—the users can choose only data introduced by validated users, who possess a certain level of trust and are less likely to introduce incorrect or inaccurate data. In case if some users intentionally introduce false data, their status, after warnings, can be lowered and their data deleted.

*Data Moderation and Approval.* ToxAlerts takes advantage of the data moderation framework inherited from OCHEM. Users can introduce new endpoints and upload structural alerts that will be immediately available to the community. However, the new data are marked as "awaiting approval" until the moderator approves the submitted data. The other users can opt to use only the approved alerts for their screening purposes. The moderators team can approve the newly entered data, correct it if necessary, request additional information, and send a message to the contributors in there are any discrepancies, mistakes, or potential inaccuracies.

*Obligatory Specification of Source.* ToxAlerts inherited from OCHEM a very important policy—the obligatory specification of the data source, e.g. citing a scientific article or a book chapter. It is not allowed to upload structural alerts with no literature source specified. In case if an unverifiable or unreliable source is specified, such data will not be approved by moderators and will not be exposed to the scientific community.

## ■ USE CASE: VIRTUAL SCREENING OF CHEMICAL LIBRARIES

To demonstrate a potential use of the ToxAlerts platform, we virtually screened screening libraries from three providers of chemicals: ChemDiv (www.chemdiv.com), LifeChemicals (www.lifechemicals.com), and Enamine (www.enamine.net) as well as the DrugBank database (www.drugbank.ca).[19] All the three suppliers provided SD files for the compound libraries and sets of structural alerts used for their internal compound

selection purposes. Therefore, we decided to virtually screen each compound library against each set of structural alerts.

**Compound Libraries.** The Enamine compound library contained almost 230 000 compounds. Enamine describes the set as "the most innovative compounds that were synthesized at Enamine within year 2010 and that emerged from more than a decade of scientific research at Enamine".[20] As claimed by Enamine, this collection features compounds with improved ADMET profiles.

Life Chemicals compound library contained more than 340 000 compounds synthesized by the company and built around 2181 original templates.[21]

The largest compound library was from ChemDiv and contained almost 400 000 chemical structures available in stock from this company. The profile of the structures from this library was similar to that of Enamine and Life Chemicals.

Additionally to the three aforementioned chemical providers libraries, we have screened the Canada DrugBank database containing 6239 structures of approved marketed and investigational small molecule drugs.[19]

**Sets of Structural Alerts.** All three commercial sets include filters based on functional groups or molecular moieties related to the covalent modification of cellular nucleophiles and proteins. Since formation of a covalent adduct is described as an initial event for toxicity,[22] the flagged compounds can cause adverse side effects.

The filters identify the compounds with electrophilic moieties (readily or upon undergoing a metabolic activation). Such compounds are capable to attack nucleophilic regions in biological macromolecules (for example, amine and thiol groups of lysine and cysteine residues in proteins) by such reactions as acylation (e.g., acid halides, ketenes, cyclic anhydrides), Michael addition (e.g., $\alpha$, $\beta$-unsaturated aldehydes, ketones, esters, amides), Schiff-base formation (e.g., form-aldehyde, aromatic aldehydes), nucleophilic aromatic substitution (e.g., nitro-substituted haloaromatics), bimolecular nucleophilic aliphatic substitution (e.g., haloalkanes, alkyl sulfates, alkyl sulfonates).

Additionally, all the structural alerts include filters to eliminate compounds like those with large planar cores (e.g., polycyclic aromatic compounds) that can intercalate with DNA, containing two or more nitro or nitrile groups, potentially unstable compounds at acidic conditions, etc. Enamine and Life Chemicals companies flag crown ethers because of their high affinity for certain metal cations (e.g., potassium or iron ions), which contributes to their toxicity as potential metal chelators.

For more detailed information on the utilized structural alert sets, refer to http://ochem.eu/alerts.

The alerts are summarized in Table 1.

**Screening Results.** The screening results for all three compound libraries as well as DrugBank against the three sets of structural alerts are summarized in Table 2.

Among the commercial suppliers, the Life Chemicals compound library had the highest number of compounds

**Table 1. Screened Libraries and Set of Structural Alerts from Three Different Providers**

| provider | alerts | screening compounds |
| --- | --- | --- |
| ChemDiv alerts | 100 | 391145 |
| Life Chemicals alerts | 72 | 343949 |
| Enamine alerts | 108 | 228899 |

**Table 2. Summary of the Screening Results for Three Compound Libraries and DrugBank against Three Sets of Structural Alerts[a]**

| | ChemDiv compounds | | Life Chemicals compounds | | Enamine compounds | | DrugBank compounds | |
|---|---|---|---|---|---|---|---|---|
| alerts set | total hits | percentage | total hits | percentage | total hits | percentage | total hits | percentage |
| all compounds | 391,145 | 100% | 343,949 | 100% | 228,899 | 100% | 6,239 | 100% |
| ChemDiv alerts | 29,430 | 8% | 39,743 | 12% | 23,616 | 10% | 1,407 | 23% |
| Life Chemicals alerts | 18,784 | 5% | 48,037 | 14% | 26,415 | 12% | 2,735 | 44% |
| Enamine alerts | 37,963 | 10% | 62,532 | 18% | 33,171 | 14% | 2,325 | 37% |
| all alerts | 71,502 | 18% | 99,929 | 29% | 53,385 | 23% | 3,697 | 59% |

[a]A "hit" here is a match of a chemical compound against at least one structural alert. The "All Sets of Structural Alerts" set flags compounds that matched at least one of the alerts in any of the sets.

with alerts. This was observed for each set of structural alerts as well as for the total number of compounds that matched at least one alert. Namely, 29% of the LifeChemicals compounds matched at least one alert in any of the three sets. For ChemDiv and Enamine libraries, these percentages were 18% and 23%, respectively.

An interesting result was observed for the DrugBank compounds of approved, discontinued, and investigational drugs. The percentage of alert-matching compounds was 2−3 times higher than in the screening compound libraries from the chemical providers. This result may seem contradictory, since the existent drugs should be relatively "safe".

It is worth stressing that the presence of a structural alert does not always translate into formation of a reactive metabolite and since the efficacious doses of newly developed investigational drugs tend to get lower and lower, many drugs even if they form a reactive metabolite provide a sufficient safety margin. In addition, a lot of drugs are taken only as needed for a short period of time. As a result, drugs get cleared rapidly from the body and the reactive metabolite (in low doses) could be neutralized by formation of a glutathione conjugate and quickly cleared. Since the safety bar for investigational new drugs gets higher and higher every year, the chemical supplier companies tend to create and sell cleaner screening compound libraries, free from the structural alerts that could cause severe adverse side effects as much as possible. This explains the fact that the DrugBank database contains a higher (percentagewise) fraction of compounds with structural alerts compared to the screening libraries from ChemDiv, Enamine, and LifeChemicals that we used in this study.

It is also very important to notice that most ADRs are not caused by the drugs themselves but by their reactive metabolites (RMs) that are formed upon bioactivation, e.g., oxidation by Cytochromes P450 in the liver or in the intestinal wall. These toxicological effects are dose-dependent and usually reversible after withdrawal of the drug dosing. Acetaminophen (also known as Tylenol in US and Doliprane or Efferalgan in Europe) has been known for a while for its liver toxicity caused by its RM formation upon metabolic oxidation of its hydroquinone-like structure forming a hepatotoxic N-acetyl-p-quinone intermediate (see Figure 3).[23] This RM reacts with glutathione (GSH, an important cellular antioxidant) or, after depleting the cellular GSH supplies, with other essential cellular nucleophiles (like alkylating Cys residues in liver enzymes) that can lead to liver damage or even death.

Moreover, the world's best selling drug Lipitor (atorvastatin calcium) has a similar mechanism of toxicity due to the presence of the structural alert aniline and does form a reactive metabolite. However, as it was reported, it is perfectly safe at doses up to 20 mg (a human dose for a person of 70 kg).[24]
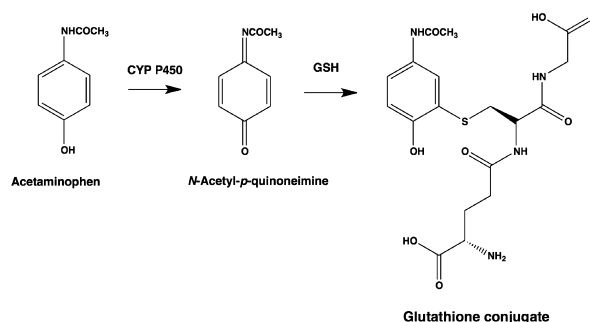


**Figure 3.** Formation of a toxic reactive metabolite from Acetaminophen.

Thus, when applying structural alerts to virtually screen chemical libraries, one should also take into account a number of other factors discussed in this article. By itself, the presence of an alert is not a sufficient justification to exclude a given structure from further development.

## IMPLEMENTATION ASPECTS

ToxAlerts extends the OCHEM framework and was built on the same technologies: Java 6, Tomcat Web server, Spring MVC framework (http://www.springsource.org), MySQL database, Hibernate ORM (http://www.hibernate.org), XSLT transformations, jQuery (http//www.jquery.com). For SMARTS parsing and matching, the MolSearch utility from ChemAxon was used (http//www.chemaxon.com).

ToxAlerts takes advantage of the distributed calculation system provided by OCHEM. For example, the screening of more than 200 000 Enamine compounds against more than 600 alerts required in total about 26 core-hours but was effectively completed in less than 50 min.

## CONCLUSIONS

ToxAlerts is a unique knowledge-based expert Web-based platform for storing structural alerts for toxic chemicals and compounds that could cause adverse drug reactions. The platform allows virtual screening of chemical compound libraries to flag potentially toxic chemicals, reactive and potentially unstable compounds, and compounds with liabilities that could form reactive metabolites upon bioactivation. Currently, the database contains more than 600 alerts from 12 publications for carcinogenicity, mutagenicity, skin sensitization, acute aquatic toxicity, and potential idiosyncratic drug toxicity.

The system is accessible from any modern Web-browser that supports Java. The platform is designed to promote collaborations and facilitate drug discovery efforts in the

2314

dx.doi.org/10.1021/ci300245q | J. Chem. Inf. Model. 2012, 52, 2310−2316

academic sector: registered users can add their own structural alerts as well as take advantage of the alerts that are already in the database to virtually screen their libraries.

The ToxAlerts system is integrated with the Online Chemical Modeling Environment (OCHEM),[18] which provides a number of benefits such as a possibility to screen against structural alerts, usage of already built and creation of new QSAR models on our parallel distributed computational environment. ToxAlerts is publicly accessible after a simple online registration procedure at http://ochem.eu/alerts or from a general OCHEM interface at http://ochem.eu.

In addition to the identification of potentially toxic compounds, the structural alerts can be also used to define structurally similar compounds that can be used for the development of "local" QSAR models for specific chemical classes of molecules. In this way, a set of structural alerts can be used to define an applicability domain of a given QSAR model,[25] which falls well within the context of the European program on Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH).

A big advantage of the structural alerts approach featured by ToxAlerts is the clear interpretability of results. While the conventional QSAR models based on many molecular descriptors and nonlinear methods can be superior for catching complex dependencies, the structural alerts is an excellent technique to flag potentially toxic compounds and provide easily interpretable results. The interpretability of predictions would also be advantageous for legislative purposes like the REACH initiative.[26]

At the moment, the coverage of structural alerts by ToxAlerts cannot compete with commercial software packages such as DEREK or MultiCASE. However, this open collaborative platform has a great potential, proven by the success of OCHEM,[18] which already contains more than 1 M data records, thousands of private models, and dozens of publicly available QSAR models. Both OCHEM and ToxAlerts are inspired by the developments in bioinformatics and biocomputing where the majority of the code and resources are publicly available and used by hundreds thousands of users. The development of open access software, depositories of data and models, and availability online resources are going to dramatically change to the way the academic community will perform chemoinfomatics and computational chemistry calculations in the future.[29]

## ■ AUTHOR INFORMATION

**Corresponding Author**
*Tel.: +49-89-3187 3575. Fax: +49-89-3187 3585. E-mail: itetko@vcclab.org.

**Notes**
The authors declare no competing financial interest.

## ■ REFERENCES

(1) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Cherkasov, A.; Li, J.; Gramatica, P.; Hansen, K.; Schroeter, T.; Müller, K.-R.; Xi, L.; Liu, H.; Yao, X.; Öberg, T.; Hormozdiari, F.; Dao, P.; Sahinalp, C.; Todeschini, R.; Polishchuk, P.; Artemenko, A.; Kuz'min, V.; Martin, T. M.; Young, D. M.; Fourches, D.; Muratov, E.; Tropsha, A.; Baskin, I.; Horvath, D.; Marcou, G.; Varnek, A.; Prokopenko, V. V.; Tetko, I. V. Applicability Domains for Classification Problems: Benchmarking of Distance to Models for Ames Mutagenicity Set. *J. Chem. Inf. Mod.* **2010**, *27*, 2094−111.
(2) Guidance on information requirements and chemical safety assessment. Chapter R.7A−Endpoint specific guidance. ECHA. http://echa.europa.eu/web/guest/guidance-documents/guidance-on-information-requirements-and-chemical-safety-assessment, (accessed July 1, 2012).
(3) Saiakhov, R. D.; Klopman, G. MultiCASE Expert Systems and the REACH Initiative. *Toxicol. Mech. Methods* **2008**, *18*, 159−175.
(4) Sanderson, D. M.; Earnshaw, C. G. Computer prediction of possible toxic action from chemical structure; the DEREK system. *Hum. Exp. Toxicol.* **1991**, *10*, 261−273.
(5) Enoch, S. J.; Ellison, C. M.; Schultz, T. W.; Cronin, M. T. D. A review of the electrophilic reaction chemistry involved in covalent protein binding relevant to toxicity. *Crit. Rev. Toxicol.* **2011**, *41*, 783−802.
(6) Verhaar, H. J. M.; vanLeeuwen, C. J.; Hermens, J. L. M. Classifying environmental pollutants. 1: Structure-activity relationships for prediction of aquatic toxicity. *Chemosphere* **1992**, *25*, 471−491.
(7) Hermens, J. L. Electrophiles and acute toxicity to fish. *Environ. Health Perspect.* **1990**, *87*, 219−225.
(8) Enoch, S. J.; Madden, J. C.; Cronin, M. T. D. Identification of mechanisms of toxic action for skin sensitisation using a SMARTS pattern based approach. *SAR QSAR Environ. Res.* **2008**, *19*, 555−578.
(9) Ashby, J.; Tennant, R. W. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat. Res.* **1988**, *204*, 17−115.
(10) Benigni, R; Bossa, C. Structure alerts for carcinogenicity, and the Salmonella assay system: a novel insight through the chemical relational databases technology. *Mutat. Res.* **2008**, *659*, 248−261.
(11) Payne, M. P.; Walsh, P. T. Structure-activity relationships for skin sensitization potential: development of structural alerts for use in knowledge-based toxicity prediction systems. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 154−161.
(12) Gerner, I.; Barratt, M. D.; Zinke, S.; Schlegel, K.; Schlede, E. Development and prevalidation of a list of structure-activity relationship rules to be used in expert systems for prediction of the skin-sensitising properties of chemicals. *Altern. Lab. Anim.* **2004**, *32*, 487−509.
(13) Bailey, A. B.; Chanderbhan, R.; Collazo-Braier, N.; Cheeseman, M. A.; Twaroski, M. L. The use of structure-activity relationship analysis in the food contact notification program. *Regul. Toxicol. Pharmacol.* **2005**, *42*, 225−235.
(14) Barratt, M. D.; Basketter, D. A.; Chamberlain, M.; Admans, G. D.; Langowski, J. J. An expert system rulebase for identifying contact allergens. *Toxicol. In Vitro.* **1994**, *8*, 1053−1060.
(15) Kazius, J.; McGuire, R.; Bursi, R. Derivation and validation of toxicophores for mutagenicity prediction. *J. Med. Chem.* **2005**, *48*, 312−320.
(16) Williams, D. P. Toxicophores: investigations in drug safety. *Toxicology* **2006**, *226*, 1−11.
(17) Chemaxon, JCHEM suite. http://www.chemaxon.com/jchem (accessed July 1, 2012).
(18) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. V.; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-de-Sousa, J.; Zhang, Q.-Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. Online chemical

2315

dx.doi.org/10.1021/ci300245q | *J. Chem. Inf. Model.* 2012, 52, 2310−2316

modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* **2011**, *25*, 533−554.

(19) Wishart, D. S.; Knox, C.; Guo, A. C.; Shrivastava, S.; Hassanali, M.; Stothard, P.; Chang, Z.; Woolsey, J. DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **2006**, *34*, D668−672.

(20) Screening Compounds Download Page−Enamine. http://www.enamine.com, (accessed July 1, 2012).

(21) Life Chemicals Collection of proprietary screening compounds. http://www.lifechemicals.com, (accessed July 1, 2012).

(22) Schwöbel, J. A. H.; Koleva, Y. K.; Enoch, S. J.; Bajot, F.; Hewitt, M.; Madden, J. C.; Roberts, D. W.; Schultz, T. W.; Cronin, M. T. D. Measurement and estimation of electrophilic reactivity for predictive toxicology. *Chem. Rev.* **2011**, *111*, 2562−2596.

(23) Kalgutkar, A. S.; Soglia, J. R. Minimising the potential for metabolic activation in drug discovery. *Expert Opin. Drug Metab. Toxicol.* **2005**, *1*, 91−142.

(24) Stepan, A. F.; Walker, D. P.; Bauman, J.; Price, D. A.; Baillie, T. A.; Kalgutkar, A. S.; Aleo, M. D. Structural alert/reactive metabolite concept as applied in medicinal chemistry to mitigate the risk of idiosyncratic drug toxicity: a perspective based on the critical examination of trends in the top 200 drugs marketed in the United States. *Chem. Res. Toxicol.* **2011**, *24*, 1345−1410.

(25) Tetko, I. V.; Bruneau, P.; Mewes, H.-W.; Rohrer, D. C.; Poda, G. I. Can we estimate the accuracy of ADME-Tox predictions? *Drug Discov. Today* **2006**, *11*, 700−707.

(26) Artemenko, N. V.; Baskin, I. I.; Palyulin, V. A.; Zefirov, N. S. Prediction of Physical Properties of Organic Compounds Using Artificial Neural Networks within the Substructure Approach. *Dokl. Chem.* **2001**, *381*, 317−320.

(27) Veith, G. D.; Call, D. J.; Brooke, L. T. Structure−Toxicity Relationships for the Fathead Minnow, Pimephales promelas: Narcotic Industrial Chemicals. *Can. J. Fish. Aquat. Sci.* **1983**, *40*, 743−748.

(28) Veith, G. D.; Broderius, S. J. Rules for distinguishing toxicants that cause type I and type II narcosis syndromes. *Environ. Health Perspect.* **1990**, *87*, 207−211.

(29) Tetko, I. V. The perspectives of computational chemistry modeling. *J. Comput. Aided. Mol. Des.* **2012**, *26*, 135−136.