Arch Toxicol
DOI 10.1007/s00204-015-1587-5

REVIEW ARTICLE

Toxicophore exploration as a screening technology for drug design and discovery: techniques, scope and limitations

Pankaj Kumar Singh[1] · Arvind Negi[1] · Pawan Kumar Gupta[2] · Monika Chauhan[1] · Raj Kumar[1]

Received: 14 July 2015 / Accepted: 13 August 2015

Abstract Toxicity is a common drawback of newly designed chemotherapeutic agents. With the exception of pharmacophore-induced toxicity (lack of selectivity at higher concentrations of a drug), the toxicity due to chemotherapeutic agents is based on the toxicophore moiety present in the drug. To date, methodologies implemented to determine toxicophores may be broadly classified into biological, bioanalytical and computational approaches. The biological approach involves analysis of bioactivated metabolites, whereas the computational approach involves a QSAR-based method, mapping techniques, an inverse docking technique and a few toxicophore identification/estimation tools. Being one of the major steps in drug discovery process, toxicophore identification has proven to be an essential screening step in drug design and development. The paper is first of its kind, attempting to cover and compare different methodologies employed in predicting and determining toxicophores with an emphasis on their scope and limitations. Such information may prove vital in the appropriate selection of methodology and can be used as screening technology by researchers to discover the toxicophoric potentials of their designed and synthesized moieties. Additionally, it can be utilized in the manipulation of molecules containing toxicophores in such a manner that their toxicities might be eliminated or removed.

Keywords Toxicophore · Screening techniques · Biological · Computational · Toxicity · Hepatotoxicity

Introduction

In recent years, adverse drug reactions (ADRs) and drug-related toxicities are the major concerns for the pharmaceutical industry involved in design, synthesis and development of a new chemical entity. In simple terms, ADR indicates the unwanted, obnoxious or harmful reactions experienced under normal conditions of use (Lee 2006). These can be a direct or indirect extension of pharmacological activity as shown in online resource 1. Direct extensions include the metabolic bioactivation of structural features, which plays a key role in initiating toxicity, known as toxicophore-induced toxicity. Indirect extensions involve toxicity related to the effects induced by the elevated concentrations of the drug. Off-target effects can be understood in terms of the similar structural resemblance of other proteins with respect to the original target; this is termed pharmacophore-induced toxicity (Sharma et al. 2011).

"Toxicophore" and "pharmacophore" are related terms, applied to biophores that are connected with some toxic and therapeutic endpoint, respectively (Richard et al. 2006). A biophore is a structural fragment that is statistically related to biological activity. A toxicophore is a qualitative structural feature that is thought to be responsible for a drug's toxic properties, as either a direct or indirect extension of pharmacological activity. However, a pharmacophore, a

Electronic supplementary material The online version of this article (doi:10.1007/s00204-015-1587-5) contains supplementary material, which is available to authorized users.

✉ Raj Kumar
raj.khunger@gmail.com; rajcps@cup.ac.in

[1] Laboratory for Drug Design and Synthesis, Centre for Pharmaceutical Sciences and Natural Products, School of Basic and Applied Sciences, Central University of Punjab, Bathinda 151 001, India

[2] Centre for Computational Sciences, School of Basic and Applied Sciences, Central University of Punjab, Bathinda 151 001, India

Published online: 04 September 2015

term introduced by Ehrlich (1909), refers to a substructure that carries the essential features responsible for a drug's pharmacological activity. The only basic difference between the two lies in their biological endpoints (Yang 2010). Figure 1 shows examples of toxicophores that have been responsible for the withdrawal of well-known drugs from the market (Stepan et al. 2011; Kalgutkar et al. 2012).

In 2008, Kortagere et al. reported the interaction of halogen compounds (toxicophores) with different enzymes (Fig. 2; Kortagere et al. 2008).

Recent advancements in the detection of toxicophores are mainly due to computational methodologies, but the biological approach is more widely used. This biological approach predominantly depends on the chemistry of the molecule that turns into a toxicophore after biotransformation (with the exception of a few cases). Currently, early-stage development in drug discovery projects employs studying of metabolites via several techniques such as gas chromatography with mass spectrometry (GC–MS), liquid chromatography with electrochemical detection (LC–EC array), LC–MS and nuclear magnetic resonance spectroscopy (NMR; Gamache et al. 2004; Goldsworthy et al. 1994].

Computational approaches have an edge over some of the shortcomings of the biological approach. Therefore, these are gaining significant interest and hence are extensively utilized currently for toxicophores and toxicity prediction. The computational methods, when coupled with mathematical and chemical–biological experimental data, have proved very useful to better understand the mechanisms through which a given chemical induces harm and ultimately, to predict adverse effects of the toxicophore. These computational approaches involve mainly quantitative structure–activity relationship (QSAR) modeling of drugs (Dudek et al. 2006). Although it is a powerful technique for the prediction of biological activity/toxicity of chemicals, the major problem is the interpretability of such models. These models are based on linear or nonlinear methods (Noorlander et al. 2008), such as multiple linear regression analysis (Xu et al. 1994), partial least square, genetic function approximation, principle component regression analysis, neural networks (Burden and Winkler 2000; Kuschewski et al. 1993) or support vector machines (Hsu and Lin 2002; Liao et al. 2007; Niazi et al. 2008), and they generally use hundreds of descriptors. Other techniques have also been found to have a wider application in toxicophore prediction and determination, such as mapping and inverse docking.

Sometimes, metabolites of drugs cause unwanted interactions (other than the desired interactions) which are responsible for toxicity, as they cause alterations in normal physiology at the cellular level; hence, these metabolites are called toxicophores. Recent research in this field disclosed that various chemical compounds that contain the same toxicophore elicit similar toxic effects. Surprisingly, it is very interesting to review how these toxicophores interact with the amino acid of the catalytic site and alter the conformation of the protein. The human body has various metabolic pathways to scavenge toxicophores such as glutathione conjugation and glucuronidation. However, their binding causes saturation of the enzyme of that specific pathway. The drugs containing toxicophores are considered safe when administered at a dose $\leq$10 mg/day because there is no saturation. Toxicophores exploration is now well accepted under certain guidelines, designating their cautious integration implies toxicophores are deliberately introduced in the structures of new chemical entities (NCE) or drugs (Williams and Park 2003).

Approaches to identify toxicophores

Biological approach

Basically, the biological approach is involved in the determination of toxicophoric moiety produced after the bioactivation/biotransformation of a drug molecule. The biological approach involves two important points: characterization of the bioactivation of a molecule and structural analysis of its toxicophoric metabolite (Munns et al. 1997). In this methodology, the molecule under investigation is radiolabeled and administered. Usually after regular intervals of monitoring, the animal is euthanized (Sherwin et al. 2003), and blood is then collected. The liver is removed, properly processed and stored at −80 °C before use. To determine the extent of covalent binding, a portion of the liver is taken, homogenized and subjected to an exhaustive solvent extraction. The amount of bioactivated molecule, bound irreversibly with the macromolecules, is determined (Dashwood 1992). Further structural elucidation of these intermediate metabolites or toxicophores can be done by different techniques such as mass spectrometry, hydrogen/deuterium exchange analysis tests and other techniques (Rufer et al. 2006).

To explore the role of mutations in the biological response to toxic agents, transgenic mutation assays are employed. Prominently, transgenic animals have been used to explore relationships between DNA adduct formation, gene mutation in target tissues and cancer (Boverhof et al. 2011). In 2013, Lu et al. used transgenic prion infected mice to study biaryl amides and hydrazones as potential therapeutics (Lu et al. 2013). To determine covalent binding of the drug to tissue and macromolecules, the availability of the radiolabeled compound is critical. Unfortunately, in few cases the presence of microsomes and nicotinamide adenine dinucleotide phosphate (NADPH) lead to bioactivation

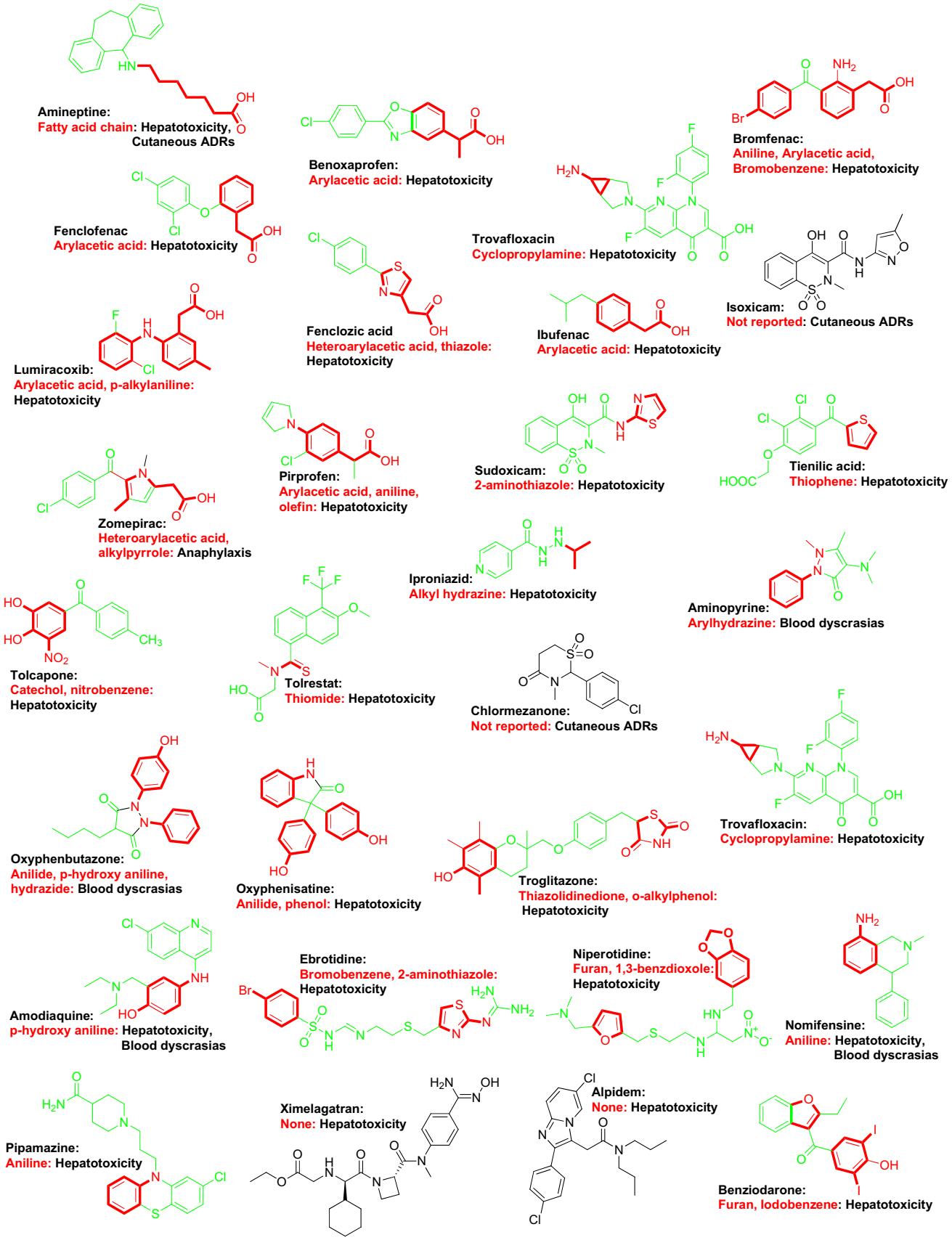
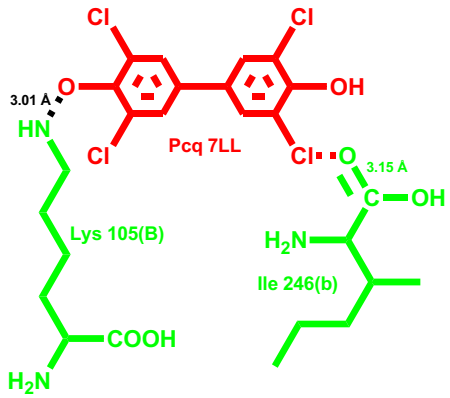
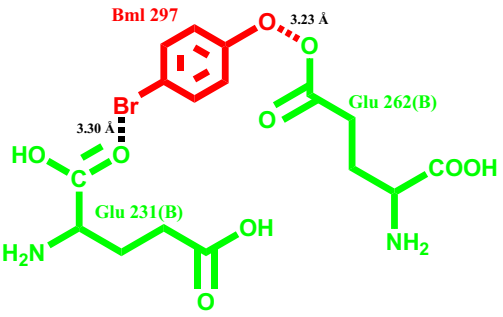


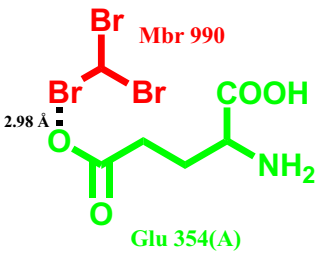
Fig. 1 Drugs withdrawn due to idiosyncratic adverse drug reactions with their toxicophores/structural alerts (shown in *red color*) (color figure online)



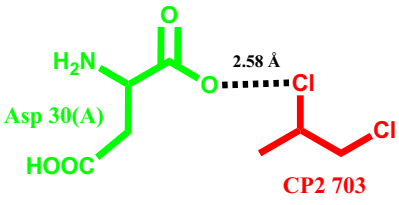
Polychlorinated biphenyl (PCB) is bound to the estrogen binding site of human estrogen sulfotransferase.



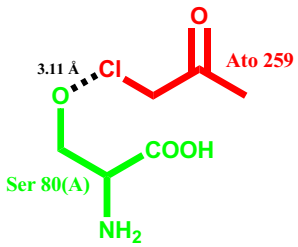
4-bromophenol bound to toluene/o-xylene monooxygenase hydroxylase.



Interaction of bromoform with firefly luciferase enzyme.



Interaction of dichloro propane with LINB, a haloalkane dehalogenase enzyme.



Interaction of chloroacetone with hydroxynitrile lyase.

Fig. 2 Interactions of few toxicophores with biomolecules (amino acids shown in *green color*; *red color* represents toxicophore) (color figure online)

of drugs that are meant to be perfectly safe in humans and act as toxicophores (Bessems and Vermeulen 2001). Protein covalent bonding is an undesirable attribute of many xenobiotics and has been responsible for their ADRs (Nakayama et al. 2009). As discussed above, these unwanted side effects can only be interpreted in terms of their metabolic activation by certain enzymes of the cytochrome P450 (CYP) superfamily. A method involving the detection of covalent modification of the CYP isoform 3A4 using MS is useful in determining the site of adduction, the nature of adducts and the extent and stoichiometry, without the need for radiolabeled compounds. This information can then be used for modification of compounds that can eliminate the possibility of interactions and covalent modification of proteins. Other techniques involved in studying metabolites include FRET, PET and NMR imaging. Methods such as PET and NMR imaging are useful for studying metabolic processes in living organisms (Zhou et al. 2012).

CYP enzymes play a dominant role in the metabolic activation of xenobiotics. They may also determine drug efficacy and bioavailability of therapeutics (Ding and Kaminsky 2003). Studies on the relationship of xenobiotic-metabolizing enzymes with the induction of toxicity in animals have been limited and difficult to interpret due to the multiple forms of expressed CYP. By introducing either the expression through genetic manipulation of the CYP enzymes in mice or the null expression of enzymes in mice, the effect of a particular enzyme in terms of chemical toxicity can be precisely determined (Jeong 1999).

For the exploration of cytotoxicity, our research group has worked on the heterocyclics as medicinal agents (Garg et al. 2015). The biological investigations have been carried out on different cell lines. The compounds have been tested on peripheral human blood lymphocytes for their toxicity evaluation (Alex et al. 2014). They were found to be non-toxic, indicating that compounds had no effect on the normal cells (Chauhan and Kumar 2013, 2014). Further various reviews covering the toxicity issues of heterocyclics belonging to different classes have been compiled (Rana et al. 2015; Chauhan and Kumar 2014; Kaur et al. 2014).

Bioanalytical techniques

Analytical techniques are implicated for accurately and selectively monitoring the multivariate analysis of endogenous metabolites in biological systems, which exhibit diverse chemical spectra over large dynamic concentration ranges.

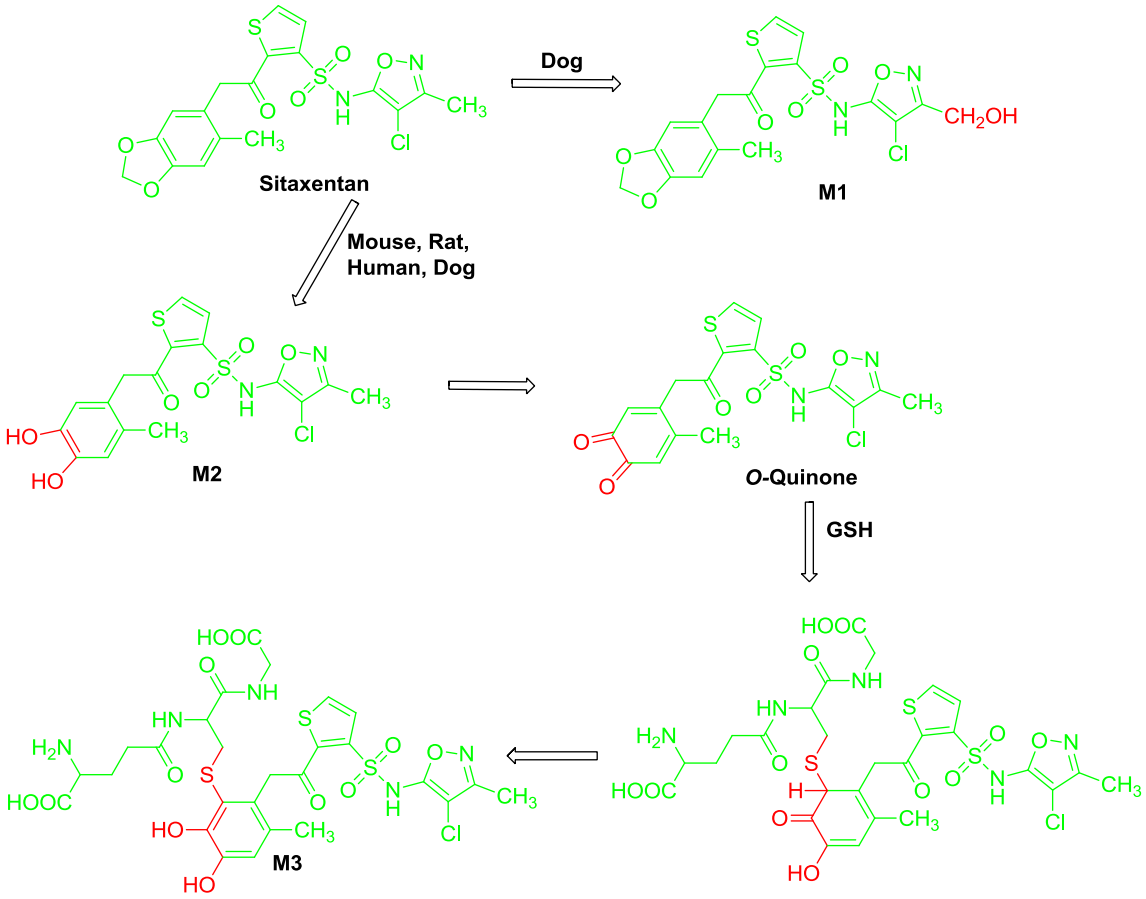


Fig. 3 Proposed metabolic pathways of sitaxentan in liver microsomes supplemented with glutathione (toxicophoric bioactivation shown in *red*) (color figure online)

These techniques are also useful for metabolomics studies, especially GC–MS, LC–EC array, LC–MS and NMR, and techniques are highly exploited (Gamache et al. 2004). These simple qualitative characterizing techniques (such as MS and NMR) provide essential data for any chromatographic peak identification and its chemical purity, for structural elucidation and for normalizing multivariate data. Although these techniques have certain limitations, NMR allows characterizing only NMR-active nuclides and also needs a comparably higher abundance of the isotope. However, MS data retrieved fragmentation pattern can be further resolved into higher resolution, which can be helpful for accurate analysis as an example of LC–MS-based metabolomics research (Schymanski et al. 2014). Moreover, the applicability of LC–EC arrays to redox-active species (e.g., hormones, neurotransmitters, antioxidants, markers of oxidative stress) has led to its increased use in the study of oxidative metabolism and redox biochemical processes, including those related to aging, immune response, inflammation and many pathological processes.

In 2013, Erve et al. studied bioactivation of sitaxentan in liver microsomes, hepatocytes and expressed human P450s, and they also characterized the glutathione conjugate by LC–MS. Sitaxentan contains a 1,3-benzodioxole ring that undergoes enzymatic demethylation to form a catechol-like metabolite that can further oxidize to a reactive *ortho*-quinone metabolite (Fig. 3).

Erve et al. (2013) reported the detection and mass spectral characterization of a glutathione conjugate of the sitaxentan quinone reactive metabolite that was trapped in vitro using mouse, rat, dog and human liver microsomes supplemented with NADPH and glutathione; this was also observed in rat and human hepatocytes. The results demonstrated that sitaxentan is capable of facile formation of a reactive *ortho*-quinone metabolite capable of reacting with glutathione and may rationalize the idiosyncratic nature of the hepatotoxicity that led to its withdrawal.

Graham et al. (2008) characterized the bioactivation of methapyrilene by hepatic microsomes and primary rat hepatocytes and established a possible causal relation with cytotoxicity. Methapyrilene (MP) tritiated at *C*-2 of the diaminoethane moiety ([3*H*] MP) was metabolized via the NADPH-dependent pathway to intermediates that irreversibly combined with microsomes (Fig. 4).

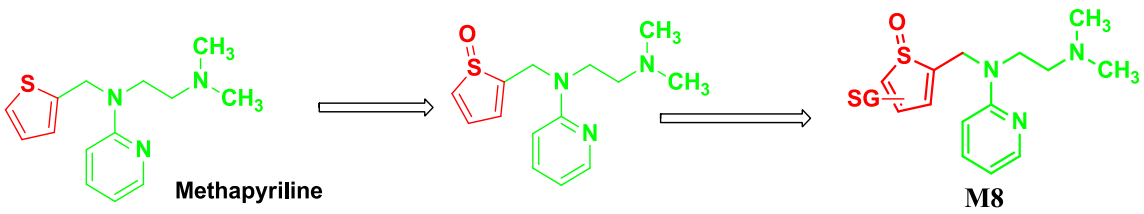


Fig. 4 Metabolic scheme for MP, showing the major stable metabolite M8 formed by primary hepatocytes from male rats. M8 is represented as a glutathione adduct derived from an S-oxide intermediate, but the position of glutathionylation was not determined (*red color* represents bioactivation) (color figure online)

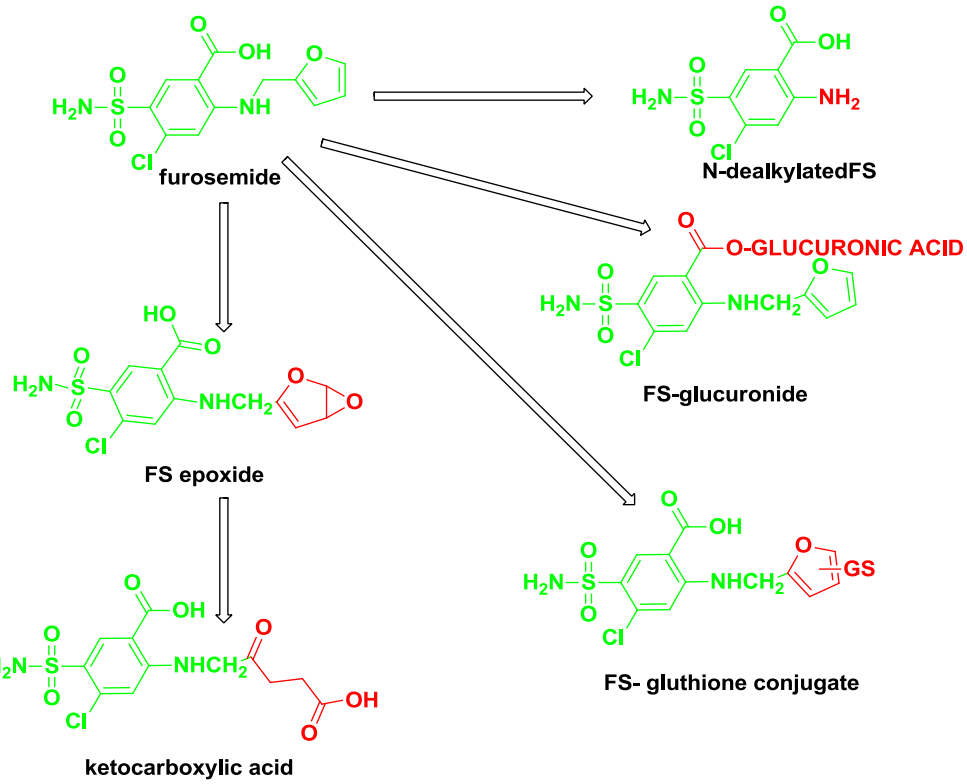


Fig. 5 In vivo metabolic pathways of furosemide in rat (bioactivation shown in *red*) (color figure online)

This binding was attenuated by the cytochrome P450 (P450) inhibitors 1-aminobenzotriazole and thiols but not by trapping agents for iminium ions and aldehydes. Reactive intermediates were trapped as thioether adducts of monooxygenated MP. Mass spectrometric and hydrogen/deuterium exchange analysis of the glutathione adduct produced by rat liver microsomes indicated that the metabolite was most likely a thioether of MP *S*-oxide substituted in the thiophene ring (Graham et al. 2008).

Williams et al. (2007) explored the nature of toxic metabolites by studying furosemide metabolism in CD1 mice and Wistar rats. It was found that furosemide at a dose of 1.21 mmol/kg caused toxicity in mice, but not in rats; this also did not result in glutathione depletion. Experiments showed that in vivo covalent binding of furosemide to hepatic proteins was sixfold higher in the mouse, i.e., 1.57 ± 0.98 nmol equivalent bound/mg protein, than in the rat, i.e., 0.26 ± 0.13 nmol equivalent bound/mg protein. However, the administration of pre-dose of the cytochrome P450 (P450) inhibitor, 1-aminobenzotriazole, reduced in vivo covalent binding to mouse hepatic proteins by 14-fold, thus reducing hepatotoxicity.

They administered [^{14}C] furosemide to bile duct-cannulated rats, which showed turnover to glutathione conjugate (8.8 ± 2.8 %), γ-ketocarboxylic acid metabolite (22.1 ± 3.3 %), *N*-dealkylated metabolite (21.1 ± 2.9 %) and furosemide glucuronide (12.8 ± 1.8 %; Fig. 5).

In mice dosed with [^{14}C] furosemide, furosemide-glutathione conjugate was not observed in bile. NMR was employed to identify the novel γ-ketocarboxylic acid, which showed the activation of furan ring (Fig. 6).

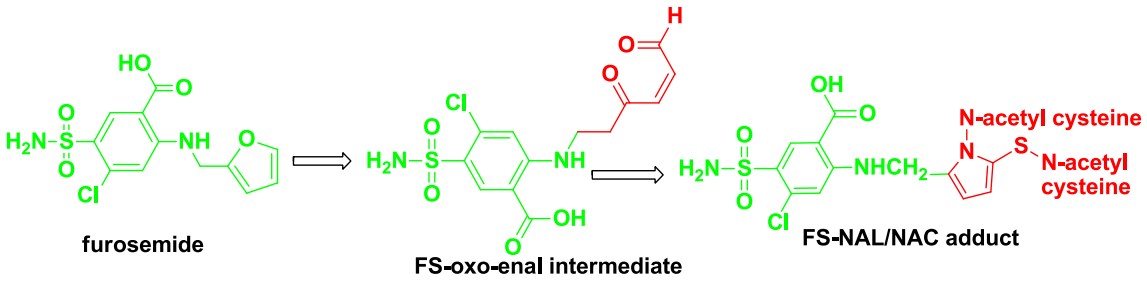


Fig. 6 In vivo metabolic pathway of furosemide in mice (bioactivation shown in *red*) (color figure online)

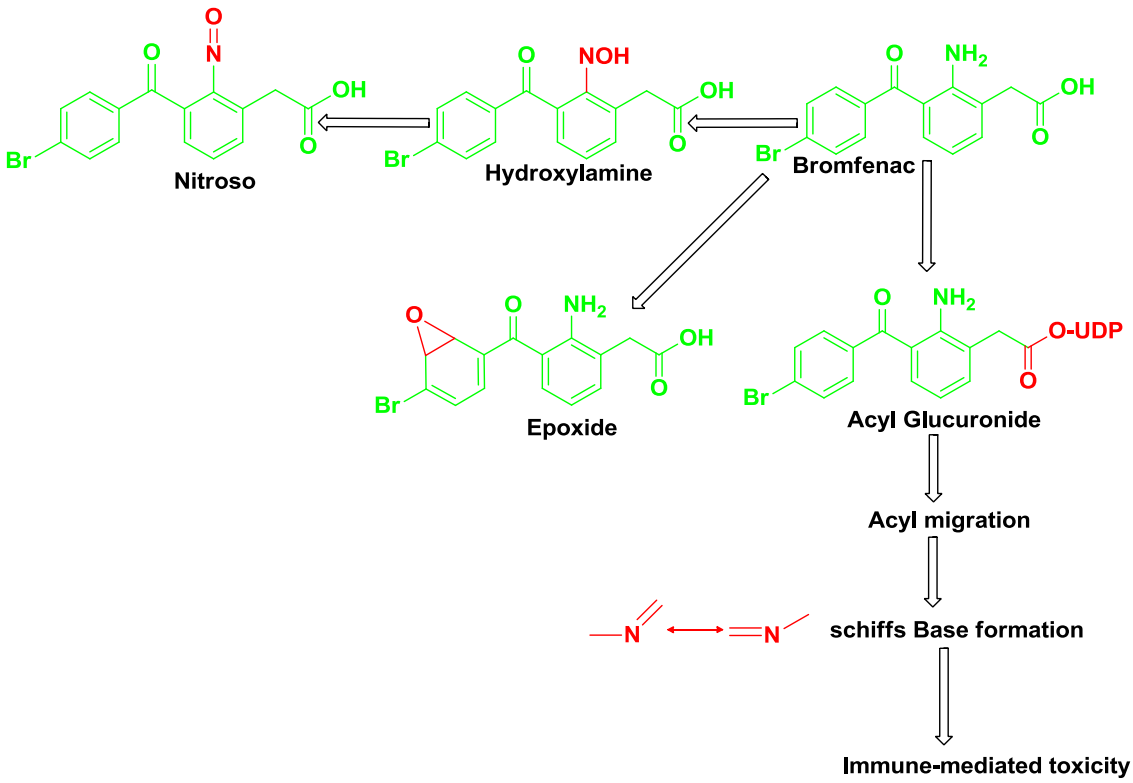


Fig. 7 Potential reactive metabolites formed from bromfenac, responsible for idiosyncratic hepatotoxicity (bioactivation shown in *red color*) (color figure online)

This also showed that formation of γ-ketocarboxylic acid was P450 dependent (Williams et al. 2007). They trapped a γ-keto-enal furosemide mouse liver microsomes, forming an *N*-acetyl cysteine/*N*-acetyl lysine furosemide adduct. They deduced that furosemide, due to bioactivation of furan ring, irreversibly bound to the rat and mouse hepatocytes. This binding significantly reduced in the presence of P450 inhibitors. Their study also underlined the need for analysis of furan rings and its activation.

Williams and Park (2003) used toxicophores present in bromfenac to explain idiosyncratic toxicity. Bromfenac was withdrawn from the market within a year of its entry. It was deduced that bromfenac consisted of three potential toxicophores. The first one is the bromophenol moiety, which is similar to hepatotoxicant bromobenzene. When the detoxification pathway via glutathione conjugation was bypassed, this metabolite caused cellular damage by binding to proteins via sulfhydryl modification. The aniline ring present in bromfenac is the second toxicophore, having the potential to form reactive nitroso compounds. This explained the unpredictable drug reactions of sulfamethoxazole, as it contains an aniline ring in its structure. The third toxicophore present in the bromfenac molecule is the arylacetic moiety (Fig. 7). Evidence suggested that phase II conjugation of arylacetic acid, 2-arylpropionic acid or anthranilic acid derivatives leads to the formation of an acyl

glucuronide, which is responsible for toxicity; however, recent studies showed that the acyl-CoA synthase-catalyzed formation of acyl-CoA thioesters has been responsible for the toxicity (Williams and Park 2003).

Computational approaches

Computational approaches have been incorporated to predict and determine toxicophore moiety (Bhavani et al. 2006; Gopi Mohan et al. 2007). Hansch and Fujita (1964) provided a fundamental scientific framework for the quantitative correlation of chemical structures with biological activity and spurred many of the developments in the field of quantitative structure–activity relationship (QSAR; Greene 2002). Frameworks that are used for toxicity and toxicophore prediction are termed quantitative structure–toxicity relationship (QSTR) models.

QSAR and QSTR work on same principle, but in QSTR, toxicity is used in place of experimental activity. The QSTR approach is used to directly predict toxicity of molecules/drugs and indirectly deduces the important physicochemical properties that are responsible for toxicity. Furthermore, these physicochemical properties can be helpful for optimization of toxicity in the molecules. Basically, the different physicochemical properties of the molecules are derived from the structural architecture of the molecules that depends on the 3-D conformation of the molecule, chemical moiety/scaffold, functional groups, aliphatic chain and/or aromatic ring. These structural features can also be interpreted as toxicophores. By changing the structural features of the molecules, different physicochemical properties can be optimized to reduce their toxicity. The optimized molecules/drugs must be verified experimentally.

On the other hand, pharmacophore/toxicophore mapping and inverse docking are also useful for prediction of the toxicophores. Recently, the reports on the use of 3-D toxicophore mapping have shown that these techniques enable us to identify essential structural attributes and to quantify the prime molecular prerequisites that are responsible for toxicity (Kar and Roy 2013). Following methodologies are employed in computational approaches:

1. Toxicophore mapping technique
2. Inverse docking approach
3. QSAR-based approach
4. Toxicophore discovery tools/toxicity predicting tools.

Toxicophore mapping technique

Toxicophore mapping is similar to pharmacophore mapping. The basic concept in toxicophore mapping is aligning the toxic molecules that have common toxicophoric features and involves utilization of software to identify possible binding features between a receptor site and a set of ligands that can explain variations in their activity. The software prepares toxicophore models based on the hypothesis that are formed based on chemical features that are important for binding to the active site. The software works on the algorithms involving these hypotheses. Generally, each hypothesis on which the software works includes four parts:

Chemical features

Functions such as hydrophobes, charged/ionizable groups and hydrogen bond donors/acceptors are considered in order to develop the model. Surface accessibility is also studied so that we only focus on the hydrophobic or hydrogen bond donor groups that are available for interaction.

Location and orientation in 3-D space

The positions of different features by absolute coordinates rather than by inter-feature distances alone should be defined. This allows discrimination between enantiomers and improves the algorithm.

Tolerance in location

Each chemical function is graphically represented by colored spheres. The size of the spheres represents the precision necessary to determine the location of a particular feature or excluded volume region. This allows to distinguish which feature location is crucial for a compound's activity.

Weight

Each chemical function includes a weight that describes its relative importance in conferring activity. The numerical value of the weight represents the order of magnitude increase in activity that can be expected from using that function fully as opposed to missing it entirely.

Feature-based alignment

Feature-based alignment of compounds is carried out without considering their activity. Further chemical features of a molecule are matched with drug molecules. Configurations having common features in a set of molecules are identified (Barnum et al. 1996). A molecule matches the configurations if it possesses conformations and structural features that can be superimposed within a certain tolerance from the corresponding ideal locations. Partial features of the molecule in the alignment set are also mapped.

Partial mapping allows us to identify larger, more diverse and more significant hypotheses and alignment of models without the risk of missing compounds that do not map to all of the toxicophore features (Smellie et al. 1995).

Software involved in toxicophore mapping are HypoGen, DISCO, Hiphop, ConFirm, etc.

In 2013, Kar and Roy developed classification and regression-based QSAR as well as three-dimensional toxicophore models for toxicity prediction of 104 organic chemicals causing bioluminescent repression of the bacterium genus Pseudomonas isolated from industrial waste water. Spatial, topological, thermodynamic, electronic, structural and E-state descriptors were used. Statistically significant and interpretable in silico models were obtained using linear discriminant analysis (classification), genetic partial least squares (regression) and 3-D toxicophore models. According to the deduced hypothesis 1 model, HYD aliphatic, HYD aliphatic and HYD aromatic arranged at specific positions formed three feature toxicophores. A triangle having a definite shape was obtained by placing these three features at the vertices. The distances between the two consecutive HYD aliphatic features were 6.515 Å, while the HYD aromatic feature was placed at a distance of 4.104 and 3.595 Å from the two hydrophobic aliphatic features. Two HYD aliphatic features and HYD aromatic feature were placed at an angle of 113.984° (Fig. 8).

The models were scrupulously validated internally as well as externally with the randomization test to prevent the possibility of chance correlation. Additionally, features such as octanol–water partition coefficients, third-order branching, CH_2 fragment or unsaturation and the presence of a higher number of electronegative atoms (specifically halogen atoms) enhanced toxicity of the chemicals (Kar and Roy 2013).

Garg et al. (2008) developed QSTR and toxicity models for a diverse series of hERG K+ channel blockers, acting as antiarrhythmic agents. A total of 68 molecules from the literature were selected having IC_{50} values measured on hERG K+ channels expressed in mammalian cell lines. Electrotopological, thermodynamic, ADMET, graph theoretical (topological and information content) descriptors were employed to derive a quantitative relationship between the channel blockers and its physicochemical properties. Statistically significant QSTR model using genetic function approximation methodology was generated, having seven descriptors. A correlation coefficient (r^2) of 0.837, a cross-validated correlation coefficient (q^2) of 0.776 and a predictive correlation coefficient (r^2 pred) of 0.701 showed the robustness of the model. Their model provided a useful framework to understand binding and gave structural insight into the specific protein–ligand interactions responsible for affinity and how one can modify any given structure to mitigate binding (Garg et al. 2008).

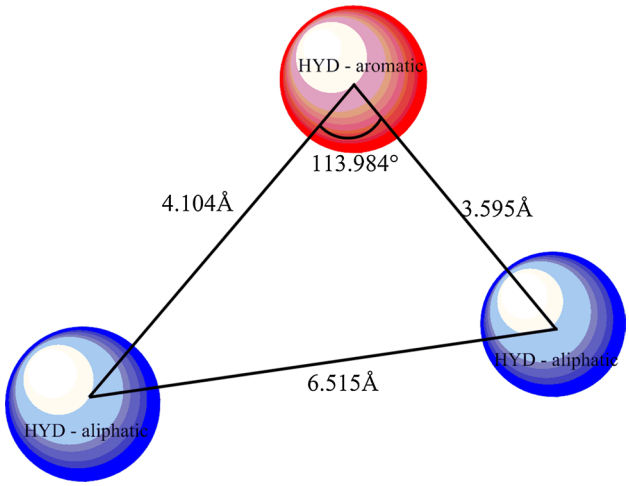


Fig. 8 Graphical representation of three feature toxicophores deduced by Kar and Roy

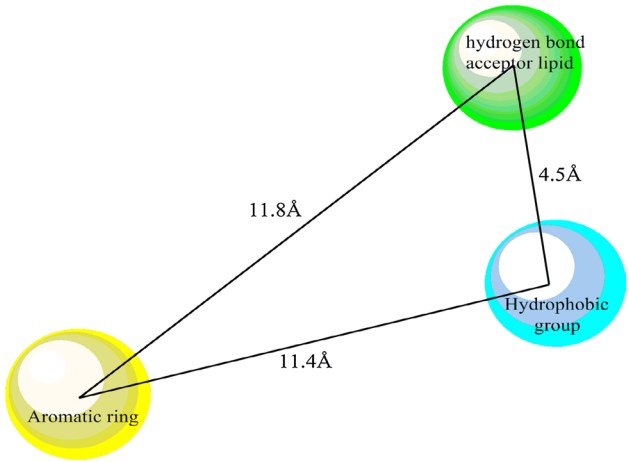


Fig. 9 Graphical representation showing three feature toxicophores deduced by Garg et al.

The best scoring HypoGen-generated toxicophore model, Hypo 1, showed three feature toxicophores having a specific arrangement of the aromatic ring, hydrophobic group and hydrogen bond acceptor lipid. The distance between the hydrophobic group and hydrogen bond acceptor lipid was 4.5 Å, the distance between the aromatic ring and the hydrophobic group was 11.4 Å, while the distance between the aromatic ring and hydrogen bond acceptor lipid was 11.8 Å (Fig. 9).

Inverse docking approach

Inverse docking approach is based on drug–receptor inverse docking methodology. In this technique, a molecule

is attempted to dock to the ligand-binding domain of proteins associated with potential toxicity and side effects. If the molecule docks well into the protein site, it is considered as toxic. The term "inverse" is used because the method is used for finding proteins that can fit a specific ligand, rather than finding ligands that fit within a specific protein (Chen and Ung 2001). The points of attachment or contact between the protein and the ligand constitute the toxicophore present in the molecule. An inverse docking procedure INVDOCK has been developed for automated search of every entry in a protein cavity database to find protein targets of a small molecule (Chen and Zhi 2001). In the INVDOCK algorithm, a drug is flexibly docked into each cavity by a procedure involving multiple conformer shape-matching alignments of the molecule to the cavity followed by molecular–mechanics torsion optimization and energy minimization on both the ligand and the binding region of the receptor. Scoring is done to identify the potential of a molecule to cause toxicity. Its main applications are (1) prediction of drug targets related to side effect and toxicity, (2) identification of unknown and secondary therapeutic targets of drugs, (3) pharmacokinetic analysis and (4) identification of unknown receptors of a ligand (Ji et al. 2006).

In 2001, Chen et al. explored the inverse docking approach to potential toxicity of small molecules having a protein cavity as targets. They selected toxicity- and side effect-related proteins and developed a protein cavity database to facilitate computer-assisted inverse docking search for target proteins. Further, they conducted an inverse docking procedure and scored based on ligand–protein interaction energy function composed of the hydrogen bond and non-bonded terms. The computer search successfully predicted 38 and missed five experimentally confirmed protein targets with an available structure in which binding involved no covalent bonds. Results on several drugs showed that 83 % of the experimentally known toxicity and side effect targets of the drugs were predicted (Chen and Ung 2001).

QSAR-based approach

QSAR is a statistically derived rule that quantitatively explains a molecular property in terms of descriptors of the chemical structure. This technique involves very complex relationships between the chemical structures and the biological properties [activity (QSAR), toxicity (QSTR) or other properties (QSPR)] and aims to deduce a relationship between the two. QSTR models are derived from the structural properties of chemical structures and experimentally determined toxicity data. Further unknown activities of the structures can be deduced by applying these models (Sanderson and Earnshaw 1991). Availability of experimental data is the limiting requirement for QSTR models. Thus, it is not possible to model every conceivable toxicity endpoint until sufficient experimental data are available.

QSTR models are required to predict the toxicity of untested compounds on the basis of existing experimental data of other compounds and to deduce the descriptors that enhance the toxicity of the molecule, which in turn can be used to determine the toxicophore. Models that are complex in terms of function complexity or the number of considered descriptors can fit almost any set of training data with high accuracy. However, such models behave poorly on future structures as they are unable to extract general relationships from the training data (high generalization error). This phenomenon, in some cases may lead to over-fitting as well as under-fitting where models fail to represent a good solution. This is because their model fitting procedure is not complex enough or because their descriptors are inadequate (Helma and Kazius 2006).

Use of a large number of irrelevant or highly correlated features can deteriorate the performance of data mining algorithms. Automated techniques exist for the removal of correlated and irrelevant features. An unsupervised method, called principal component analysis (PCA), transforms the initial set of features into a smaller, uncorrelated set of descriptor-based functions. With supervised techniques, it is possible not only to produce a smaller set of features, but also to determine the relevance of descriptors for a particular toxicity endpoint. This can be achieved with simple statistical filters that decide, for example, whether a feature occurs more frequently in toxic than in non-toxic structures. Generalized linear models use statistical regression techniques to minimize the difference between the predicted and the real values. Multiple linear regression attempts to identify a linear function that relates descriptors to toxicity values (Eriksson et al. 2005).

In 2005, Kazius et al. identified novel toxicophores using previously known "general toxicophores" as shown in online resource 2. They approved a substructure as a specific toxicophore if it simultaneously satisfied four criteria: (1) Substructure must have a sufficient degree of either intrinsic reactivity or chemical similarity with an existing knowledge-based toxicophore, or the substructure must be reported for several compounds as a critical component of a mechanism of action that leads to mutagenicity. (2) The substructure must be a toxicophore in at least three chemically different compound classes. This requirement is based on an assumption fundamental to this approach that the mutagenic character of a given substructure is generally conserved throughout the chemically diverse classes. (3) The accuracy of the substructure must at least be 70 %. (4) The substructure's p value must be smaller than 0.05. They deduced final set of 29 toxicophores containing new

Table 1 Computational predictive software tools

S. no	Softwares	URLs	Web servers	Licensed	Open access
1	DEREK	http://www.chem.leeds.ac.uk/		✓	
2	Hazard Expert	www.compudrug.com/hazard.html		✓	
3	OncoLogic	http://www.logichem.com/		✓	
4	LAZAR	http://www.predictive-toxicology.org/lazar.			✓
5	TOPKAT	www.accelrys.com		✓	
6	Protox	http://tox.charite.de/tox	✓		✓
7	COMPACT				✓
8	CASE (MCASE QSAR-ES)	http://www.multicase.com		✓	
9	MetabolExpert	www.compudrug.com		✓	

substructures as shown in online resource 3, which could classify the mutagenicity of the investigated dataset with a total classification error of 18 % (Kazius et al. 2005).

Toxicophore discovery and predicting tools

Toxicophore discovery tools (Table 1) are advanced systems that selectively predict and determine a toxicophore. They have been defined as any formal system, not necessarily computer based, that enables a user to obtain rational predictions about the toxicity of chemicals (Merlot et al. 2003).

Firstly, DEREK software got attention within the scientific community, which works on the deductive estimation of risk from existing scientific knowledge of structure–toxicity relationships and mechanisms, utilizing PATRAN language to predict the molecule toxicity (ATSDR 1997). DEREK is basically a graphical interface, along with graphical rule editor, batch processing feature, etc. On the other hand, it has certain disadvantages, as it requires elaborate information regarding activating and detoxification effects of metabolism which is not provided.

Later on, the REX tool was incorporated, which tells particularly a structural feature of the molecule which is concerned to that toxicity. REX works on two unique features. Firstly, the descriptors it uses are "atom pairs" rather than specific fragments or atom and bond chains. This feature helps find the 3-D spatial arrangement of the atoms not based on their types; this is quite significant as typical biological interactions occur with the binding centers at certain distances apart in 3-D space. A second unique feature of REX is that once the atom pairs that appear to be related to activity have been identified, they are mapped back onto the structures of the active molecules. If these overlaps are found consistently between the atoms, they are then fused together to produce a particular 3-D spatial fragment of a molecule as a toxicophore (Judson 1994).

Silva et al. predicted metabolic reactions and toxicophoric groups of psoralen and bergapten using DEREK, along with another computational metabolite-generating software METEOR. The computational analysis revealed the presence of at least six toxicophoric groups in these molecules, namely psoralen, furan, epoxide, resorcinol, coumarin and phenyl ester (Fig. 10), related to carcinogenicity, mutagenicity, photoallergenicity, hepatotoxicity and skin sensitization.

Whereas psoralen core can be considered both pharmacophoric and toxicophoric as it can cause chromosome damage and mutagenicity probably due to cross-linkage with DNA, the same mechanism was also observed for its therapeutic action. Bergapten was expected to be more efficient than psoralen because its methoxy group in the aromatic ring could favor the biotransformation at this site and minimize the epoxidation of the double bonds; therefore, the low chances of epoxy metabolite formation concurrently reduce the toxicity. Additionally, coumarin was not considered a toxic group itself but was classified as a toxicophoric marker because some of its metabolites can cause photoallergenicity (da Silva et al. 2009).

There are certainly other examples which utilized DEREK in toxicophore determination skin sensitization of citronellal, the toxicophore responsible for skin sensitization of diacetyl-diperoxyadipic acid and mutagenicity toxicophore of bisfuranoids (Fig. 11; Cronin et al. 2003).

In 1996, a research project governed by a collaboration between Ross King (University of Oxford, UK) and Mike Sternberg (Imperial Cancer Research Fund, UK) explored PROGOL use in the discovery of toxicophores and pharmacophores in sets of active chemicals. It utilizes the inductive logic and does not depend on the chemical structure descriptors. It correlates biological activity and a particular group of atoms. It has an edge over the conventional toxicity predicting software as it does not need descriptors and therefore it avoids making arbitrary decisions based on descriptor types (King et al. 1996; Ray et al. 2004).

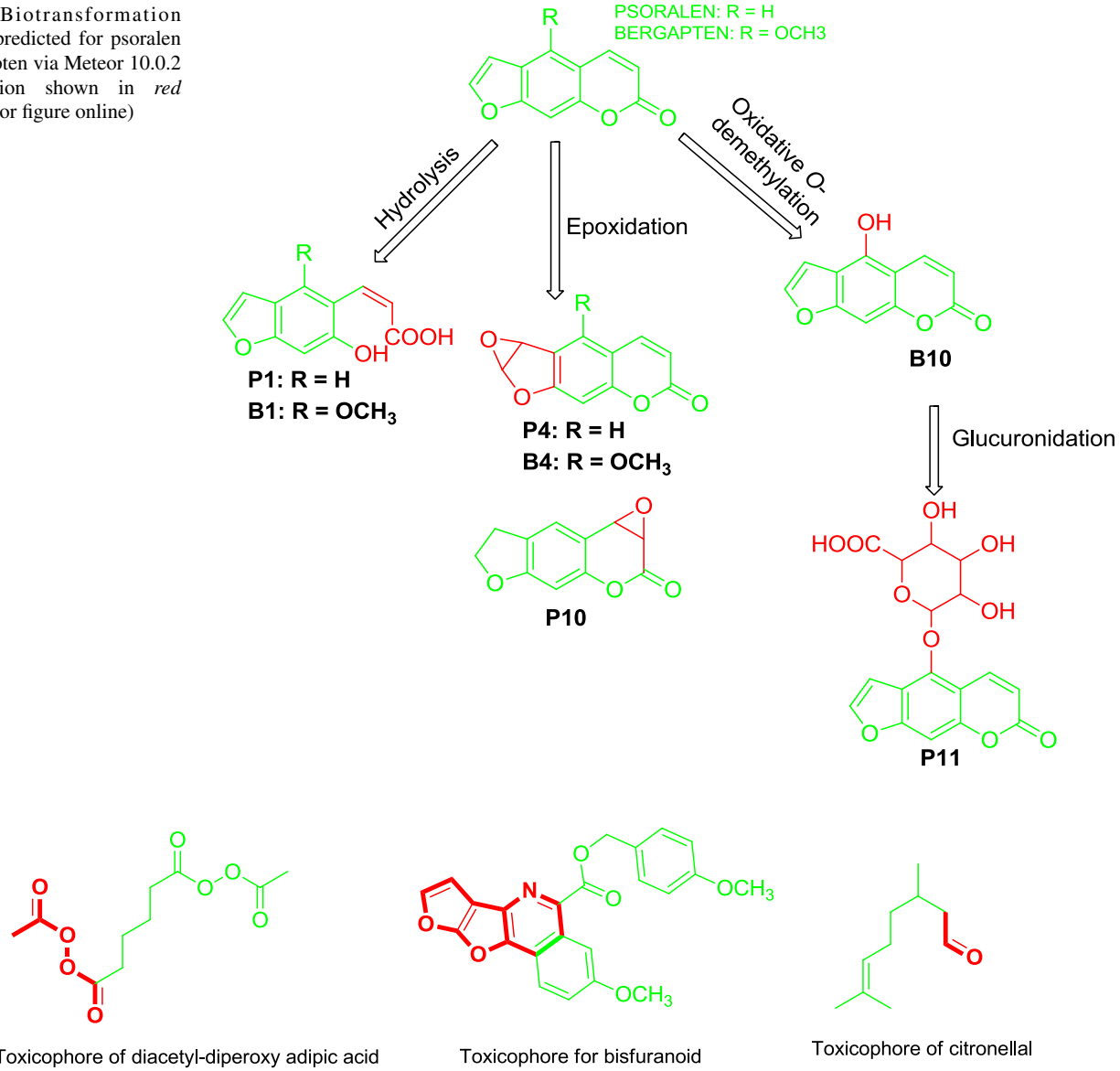


Fig. 10 Biotransformation reactions predicted for psoralen and bergapten via Meteor 10.0.2 (bioactivation shown in *red color*) (color figure online)

Fig. 11 Other toxicophores predicted via *DEREK* (*red color* represents toxicophore) (color figure online)

In 1997, DTOX was disclosed from a research project carried out for MAFF by Integral Solutions Ltd and utilizes a data mining principle in discovering the toxicophores from the provided data about the chemical structures. The main techniques involved were the induction prediction models based on an ID3-like rule induction method and neural networks. The main shortcoming of its functionality was in choosing chemical substructure descriptors and to overcome this issue; several thousand descriptors of various types were selected so that their behaviors could be compared in terms of atom and bond chain-based augmented atoms, ring indices, topological atom pairs, pairs of binding centers with 3-D through-space distances and 3-D three-center pharmacophores. Correlations between the sets of descriptors and their activity were found, but their reliability was almost found insignificant in many cases; therefore, its existence was diminished (Dearden et al. 1997).

Furthermore, TOPKAT was introduced which stands for Toxicity Prediction by Komputer-Assisted Technology. It works on an automated QSAR technique, which was designed by Health Designs along with Accelrys. TOPKAT utilizes large, heterogeneous databases with carefully selected data, and its QSARs employ mainly topological, substructural and electronic descriptors (Snyder et al. 2004). It accurately and rapidly assesses the toxicity of the chemicals solely from their two-dimensional (2D) molecular structures. It uses a range of robust, cross-validated QSTR models for assessing specific toxicological

endpoints. The software utilizes both continuous and binary measures to predict the number of toxicity endpoints, including carcinogenicity, mutagenicity, teratogenicity, irritation, allergic contact dermatitis (ACD), acute toxicity and Ah receptor binding. It also predicts the skin permeability. Continuous endpoints, such as LD_{50}, are modeled using multiple linear regression QSARs, while binary measures, such as carcinogenicity, are modeled using linear discriminant regression. Easy use, time efficiency and various toxicity prediction modules are a few merits. Moreover, TOPKAT informs the user whether the prediction falls under optimum prediction space or not. The assumption regarding the individual contribution of a substructure toward toxicity is not always correct. It also lacks the batch processing capability. Babu et al. (2014) performed in silico toxicity prediction of nilutamide (NLM) by using TOPKAT software. The toxicities of NLM and its degradation products were assessed and compared in different animal models, and both NLM and its degradation products showed toxicity and carcinogenicity. The probabilities of degradation products being toxic were higher than those of NLM.

Afterward, LAZAR, also known as lazy structure–activity relationships (ARI system), a very useful tool for prediction of the toxic properties of chemical structures, was introduced. It derives predictions for query molecules from an inductive database that contains the experimentally determined toxicity data. The predictive power of LAZAR mainly depends on the quality of data fed in its inductive database. LAZAR uses a modified *k*-nearest neighbor (kNN) algorithm for its predictions. It searches the database for a training set and its experimental data, which are similar to the query structure (neighbors), and then calculates a prediction from the experimental measurements of the query structure (Helma 2006). It directly provides applicability domain estimation, although it needs external components to properly run. Piparo et al. (2014) utilized LAZAR to describe how two models (one for the mouse and one for the rat) for the carcinogenic potency (TD_{50}) prediction have been developed.

Computer-automated structure evaluation (CASE) technology refers to a range of different programs that are supplied by MULTICASE Inc. (Cleveland, OH, USA), which are known as ToxAlert, CASE, Multi-CASE and CASETOX. It is a hybrid of 2D-QSAR and an artificial expert structure-based program. CASE technology uses a very different approach through the creation of its own structural alerts. Each molecule is broken down into the maximum number of possible fragments, usually ranging from two to ten heavy (non-hydrogen) atoms. These fragments are further classified statistically as biophores (associated with toxicity) or biophobes (not associated with toxicity) (Klopman 1992). The results are then combined to form an equation:

$$\text{CASE units} = \text{constant} + a\,(\text{fragment 1}) + b\,(\text{fragment 2}) + \cdots$$

CASE can be applied on the molecules whose mechanisms of action are not known and can also be predicted because its batch processing is very fast. However, sometimes outputs are ambiguous and can lead to misinterpretation of the predictions. Additionally, it fails to distinguish between small chains that are present in a complex and those present separately. Klopman et al. (1999) used a multiple computer-automated structure evaluation program to construct an acute fish toxicity model on the basis of a wide series of experimental data for the guppy. The created model possessed very good predictive ability. It correctly predicted acute toxicity for the guppy for 80 % of compounds with an average error of only 0.63 log unit per median lethal concentration. The importance of the necrosis effect was demonstrated. The main toxicophores, corresponding to polar necrosis and to the reactive chemicals, were identified.

OncoLogic only predicts carcinogenicity. It is a knowledge-based system developed and marketed by Logi-Chem. It uses a hierarchical, decision-tree structure for each of the four separate sub-systems for estimating the carcinogenicity of the fibers, metals and metal-containing compounds, polymers and organics, respectively (Benigni et al. 2012). Primarily, it is based on the cancer bioassay data from IARC, NCI/National Toxicology Program (NTP), and U.S. Public Health Service publication series and U.S. EPA research data. It has an expertise in the evaluation of carcinogenicity, and on the other side, it is unable to calculate or use physicochemical parameters as a part of evaluation. It is only useful in predicting carcinogenicity.

Computer-optimized molecular parametric analysis of chemical toxicity (COMPACT) is a methodology developed by Lewis et al., at the University of Surrey in the UK. COMPACT essentially predicts the potential of a chemical to act as a substrate for one of the cytochromes P450 and is based on the ability of a chemical to fit onto, and interact with, the relevant binding site on the enzyme. It relies largely on two descriptors, molecular planarity and electronic activation energy (Parke et al. 1990). It is useful in understanding P450 specificity and P450-mediated toxicity and carcinogenicity. The COMPACT radius is easy to calculate and simple to apply to the new molecules. A whole molecule approach is used rather than the fragment-based approach. It is capable of handling molecules having up to 150 (non-hydrogen) atoms. Conversely, it cannot identify the directly acting carcinogens, which do not require bioactivation via P450 (Lewis et al. 1995). Lewis et al. evaluated a series of 30 miscellaneous National Toxicology Program chemicals prospectively for carcinogenicity and toxicity by COMPACT. Evaluations were also made by HazardExpert, for metal ion redox

potentials; these, together with COMPACT, were compared with results from the Ames test for mutagenicity in Salmonella, the micronucleus test and 90-day subchronic rodent pathology. Seven of the 30 chemicals (nitromethane, chloroprene, xylenesulphonic acid, furfuryl alcohol, anthraquinone, emodin and cinnamaldehyde) were positive for potential carcinogenicity in the COMPACT evaluation; xylenesulphonic acid and furfuryl alcohol were only equivocally positive. Four of the 30 chemicals—scopolamine, D&C yellow No. 11, citral and cinnamaldehyde—were positive by HazardExpert; six of 30—D&C yellow No. 11, 1-chloro-2-propanol, anthraquinone, emodin, sodium nitrite, cinnamaldehyde—were positive in the Ames test; two of 30—phenolphthalein and emodin—were positive in the in vivo cytogenetics test; and three of 30—molybdenum trioxide, gallium arsenide and vanadium pentoxide—were metal compounds with redox potentials of the metal/metal ion indicative of possible carcinogenicity. The overall prediction for carcinogenicity was positive for 12 of 30 chemicals: nitromethane, chloroprene, D&C yellow No. 11, molybdenum trioxide, 1-chloro-2-propanol, furfuryl alcohol, gallium arsenide, anthraquinone, emodin, sodium nitrite, cinnamaldehyde and vanadium pentoxide). Overall, predictions were made on the basis of the results of the computer tests and from consideration of the information from bacterial mutagenicity, together with likely lipid solubility and pathways of metabolism and elimination (Lewis et al. 1996).

HazardExpert, a production of CompuDrugChemisty Ltd., Hungary, is another rule-based approach for the toxicity prediction (Smithing and Darvas 1992). Its functioning is based on searching the query structure for the known toxicophores that are derived from literature in the field of QSAR or from the United States EPA and Interagency Testing Committee (ITC) monographs. If a toxicophore is identified, then it triggers the estimation of number of toxicity endpoints, such as mutagenicity, carcinogenicity and teratogenicity, based on the rules in the knowledge base. The physicochemical properties such as molecular weight, log P and pKa are also calculated and used in QSAR equations.

It is also connected to MetabolExpert, a system used to determine the effect of metabolism on the query compound. It is based on the physicochemical properties for its predictions. Additionally, it provides data regarding bioavailability and bioaccumulation. The knowledge base can be edited by the user. However, it does not provide indication for the relative probabilities for formation of the metabolites (Dearden et al. 1997).

Warmr is an inductive logic programming algorithm. It uses a data log, which is a programming language specifically designed to implement deductive databases. It can discover knowledge in a structured data, where the pattern reflects one-to-many and many-to-many relationships, which is not possible with the standard data mining programs. It is based on a level-wise method from an Apriori algorithm; it performs a breadth-first search of the pattern space. This method searches for the pattern space one level at a time, starting from the most general pattern. This method iterates between two phases: candidate generation, i.e., lattice structure is used for pruning non-frequent pattern from the next level, and candidate evaluation, i.e., the frequencies of a candidate are computed with respect to the databases as shown in online resource 4 (King et al. 2001).

ProTox is a web server used for the prediction of rodent oral toxicity. This prediction method is based on the analysis of the similarity between the compounds with known median lethal doses (LD_{50}) and the incorporation of the identified toxic fragments, thereby representing a novel approach in toxicity prediction. In addition, the web server includes an indication of the total number of possible toxicity targets, which is based on an in-house collection of protein–ligand-based pharmacophore models ("toxicophores") for the targets associated with adverse drug reactions. The ProTox web server is open to all the users and can be accessed without registration at: http://tox.charite.de/tox. The only requirement for the prediction is a two-dimensional structure of the input compounds. All ProTox methods have been evaluated based on a diverse external validation set, and these displayed a strong performance (sensitivity, specificity and precision of 76, 95 and 75 %, respectively) and superiority over other toxicity prediction tools, indicating their possible applicability for other compound classes (Drwal 2014).

Conclusions

With the increasing demand of the drug safety predictions, the detection of toxicophoric moiety or substructure has become an essential step of the new screening technologies inspired by medicinal chemistry and molecular design. This concept is useful in designing a potent pharmacological agent with reduced toxicity-producing interactions. Further, the reliability and accuracy of mutagenicity, hepatotoxicity or cardiotoxicity prediction related to the drug moiety can be achieved by the analysis of toxicophore. The biological approach has emerged as the most reliable technique for studying toxicophore using trivial testing. This gives a live model which testifies the toxicity by measuring total biotransformation of the drug under test. Further, it does structural analysis of the toxicophore moiety via initially radiolabelling and further collaborating with analytical methods. The use of GC–MS, LC–MS array, NMR and H/D exchange analytical tests has benefited the biological approach. From the biological

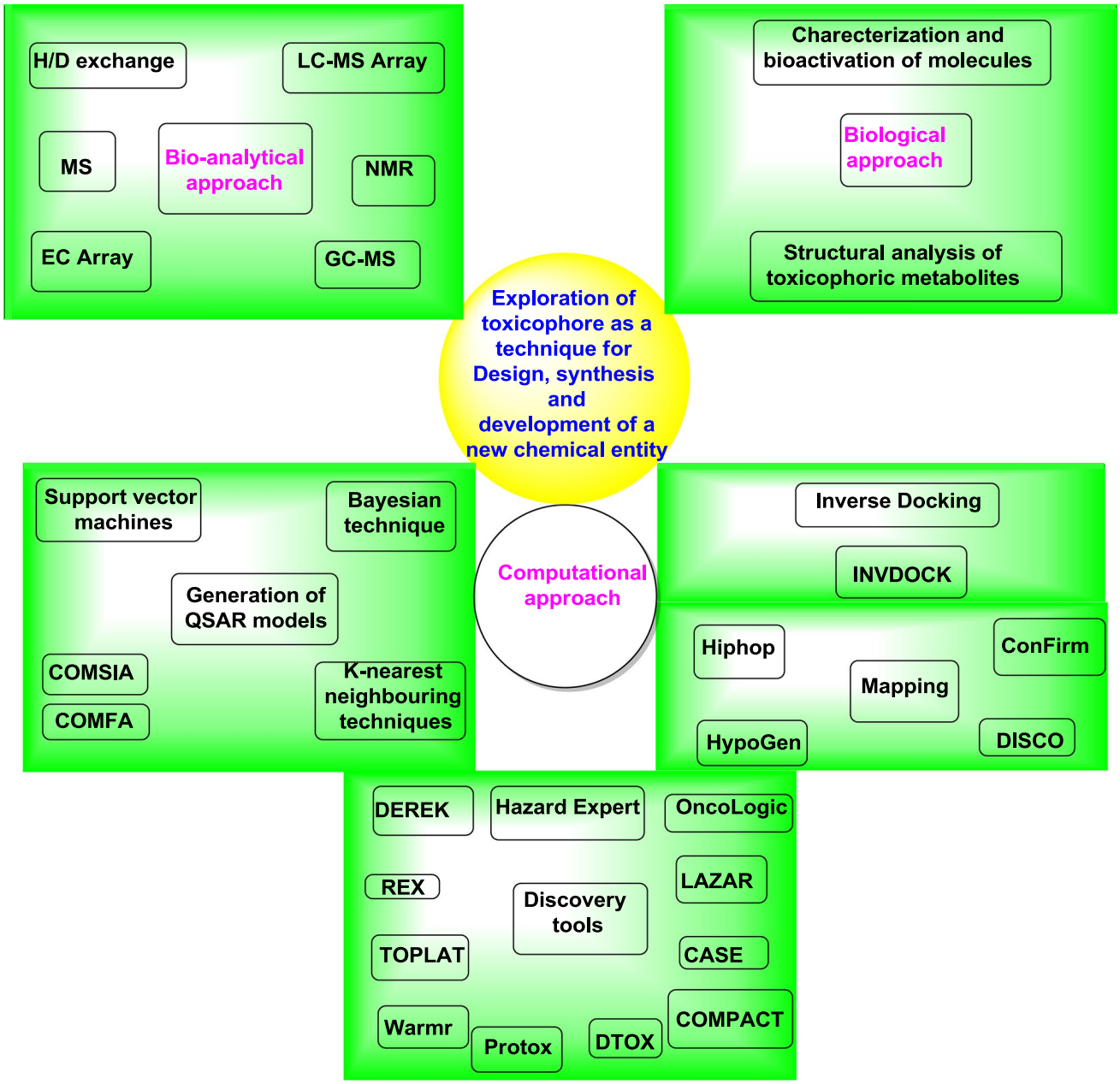


Fig. 12 Various techniques of toxicophore exploration as a screening technology for drug design and discovery

investigations to bioanalytical methods, the ethical and analytical (e.g., NMR only characterizes NMR-active nuclides) limitations led to the synchronized utilization of computational methods. The computational methods include QSTR models and prediction of physicochemical properties, binding patterns and structural features. These are used to reduce the toxic features in a molecule. Toxicophore mapping, inverse docking and toxicophore discovery tools are the various techniques used to verify results obtained from biological/bioanalytical methods. Bayesian technique, RP analogues, SVMs and FNNs have been the trivial basis for toxicophore prediction, but nowadays more advanced software such as INVDOCK and tools such as DEREK, REX, PROGOL, DTOX, TOPKAT and LAZAR are utilized (Fig. 12). Until now, there is no well-defined standard operating procedure regarding the determination of a toxicophoric moiety.

We are of the opinion that a standard protocol should comprise extraction of biological data combined with bioanalytical techniques, i.e., IC_{50} values, determination of

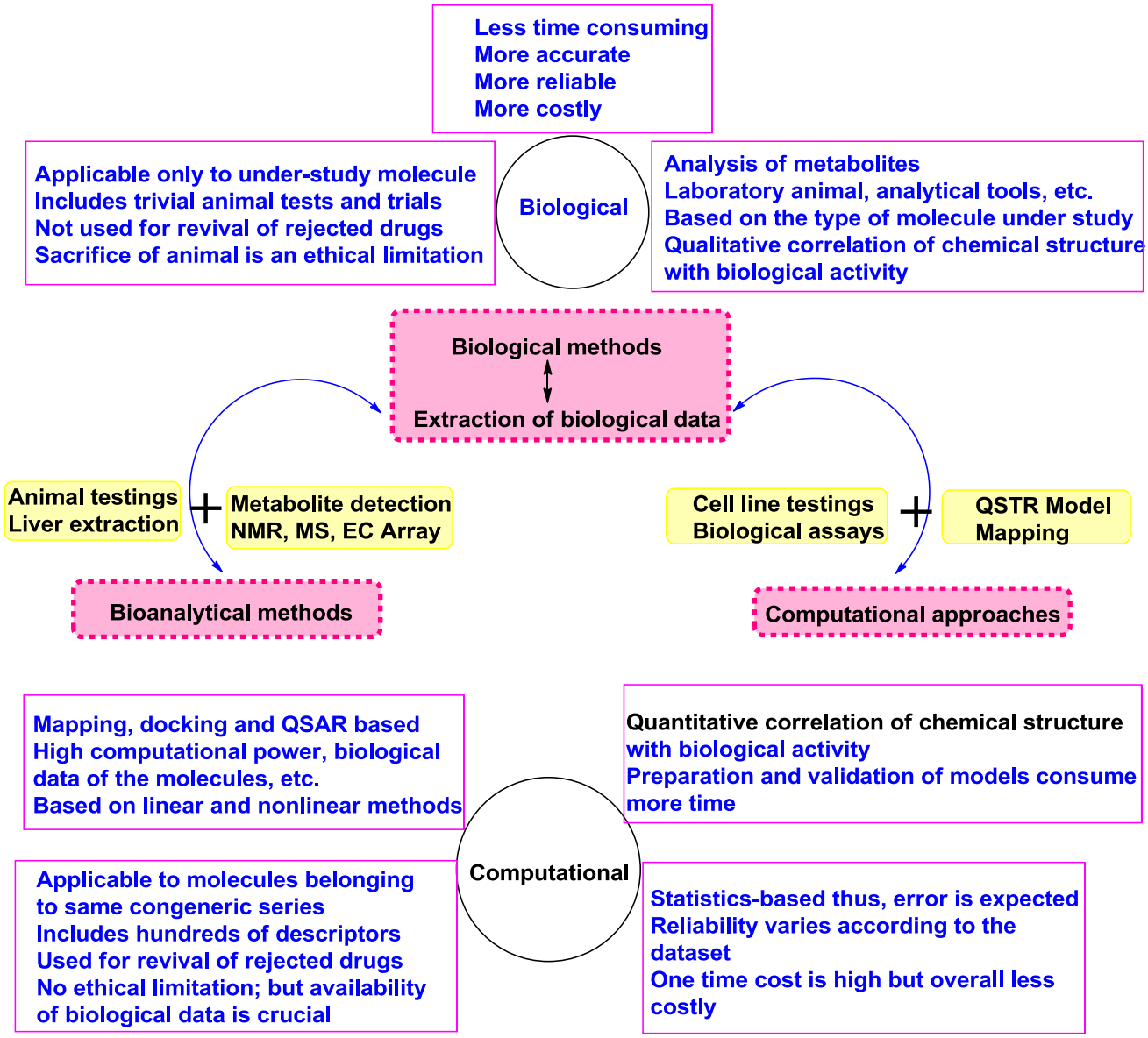


Fig. 13 Standard protocol comprising of biological, bioanalytical and computational approaches for the exploration of toxicophore

physiological and ADME data of the molecule, determination of computational data with 3-D conformations, configurations of the molecule and selection of the most relevant descriptors for the generation of the regression equation to deduce a model, which can predict the toxicophores in the new datasets (Fig. 13).

Acknowledgments R.K. and M.C. thank Department of Science and Technology, New Delhi, for the financial assistance (F. No. SR/FT/CS-71/2011). Support from Research Seed Money (RSM) from CUPB is also acknowledged. R.K. also thanks University Grant Commission-Major Grant (F. No. 42-676/2013(SR).

Compliance with ethical standards

Conflict of interest Authors do not have any conflict of interest.

References

Alex JM, Singh S, Kumar R (2014) 1-Acetyl-3, 5-diaryl-4, 5-dihydro (1H) pyrazoles: exhibiting anticancer activity through intracellular ROS scavenging and the mitochondria-dependent death pathway. Arch Pharm 347:717–727

ATSDR U (1997) Agency for toxic substances and disease registry, Case Studies in environmental medicine. http://www.atsdr.cdc.gov/HEC/CSEM/csem.html

Babu RA, Borkar RM, Raju G, Raju B, Srinivas R (2014) Liquid chromatography electrospray ionization tandem mass spectrometry study of nilutamide and its stress degradation products: in silico toxicity prediction of degradation products. Biomed Chromatogr 28:788–793

Barnum D, Greene J, Smellie A, Sprague P (1996) Identification of common functional configurations among molecules. J Chem Inf Comput Sci 36:563–571

Benigni R, Bossa C, Alivernini S, Colafranceschi M (2012) Assessment and validation of US EPA's OncoLogic® Expert system and analysis of its modulating factors for structural alerts. J Environ Sci Health 30:152–173

Bessems JG, Vermeulen NP (2001) Paracetamol (acetaminophen)-induced toxicity: molecular and biochemical mechanisms, analogues and protective approaches. CRC Crit Rev Toxicol 31:55–138

Bhavani S, Nagargadde A, Thawani A, Sridhar V, Chandra N (2006) Substructure-based support vector machine classifiers for prediction of adverse effects in diverse classes of drugs. J Chem Inf Model 46:2478–2486

Boverhof DR, Chamberlain MP, Elcombe CR, Gonzalez FJ, Heflich RH, Hernandez LG, Jacobs Jacobson-Kram D, Luijten M, Maggi A (2011) Transgenic animal models in toxicology: historical perspectives and future outlook. Toxicol Sci 121:207–233

Burden FR, Winkler DA (2000) A quantitative structure-activity relationships model for the acute toxicity of substituted benzenes to *Tetrahymena pyriformis* using Bayesian-regularized neural networks. Chem Res Toxicol 13:436–440

Chauhan M, Kumar R (2013) Medicinal attributes of pyrazolo [3, 4-*d*] pyrimidines: a review. Bioorg Med Chem 21:5657–5668

Chauhan M, Kumar R (2014) A comprehensive review on bioactive fused heterocycles as purine-utilizing enzymes inhibitors. Med Chem Res 24:2259–2282

Chen Y, Ung C (2001) Prediction of potential toxicity and side effect protein targets of a small molecule by a ligand–protein inverse docking approach. J Mol Graph Model 20:199–218

Chen Y, Zhi D (2001) Ligand–protein inverse docking and its potential use in the computer search of protein targets of a small molecule. Protein Struct Funct Bioinform 43:217–226

Cronin MT, Walker JD, Jaworska JS, Comber MH, Watts CD, Worth AP (2003) Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. Environ Health Perspect 111:376

da Silva VB, Kawano DF, Carvalho I, Conceicao EC, Freitas O, de Paula Silva CHT (2009) Psoralen and bergapten: in silico metabolism and toxicophoric analysis of drugs used to treat vitiligo. Int J Pharm Pharm Sci 12:378–387

Dashwood RH (1992) Protection by chlorophyllin against the covalent binding of 2-amino-3-methylimidazo [4, 5-*f*] quinoline (IQ) to rat liver DNA. Carcinogenesis 13:113–118

Dearden JC, Barratt MD, Benigni R, Bristol DW, Combes RD, Cronin MT, Judson PN, Payne MP, Richard AM, Tichy M (1997) The development and validation of expert systems for predicting toxicity. In: Workshop (ECVAM Workshop 24), pp 2

Ding Xinxin, Kaminsky LS (2003) Human extrahepatic cytochromes P450: function in xenobiotic metabolism and tissue-selective chemical toxicity in the respiratory and gastrointestinal tracts. Annu Rev Pharmacol Toxicol 43:149–173

Drwal MN (2014) ProTox: a web server for the in silico prediction of rodent oral toxicity. Nucleic Acids Res 42:W53–W58

Dudek AZ, Arodz T, Galvez J (2006) Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Comb Chem High Throughput Screen 9:213–228

Ehrlich P (1909) Über den jetzigen Stand der Chemotherapie. Ber Dtsch Chem Ges 42:17–47

Eriksson L, Johansson E, Lundstedt T (2005) Regression-and projection-based approaches in predictive toxicology. In: Helma C (ed) predictive toxicology, 1st edn. Taylor & Francis, New York, pp 177–221

Erve JC, Gauby S, Maynard JW Jr, Svensson MA, Tonn G, Quinn KP (2013) Bioactivation of sitaxentan in liver microsomes, hepatocytes, and expressed human P450s with characterization of the glutathione conjugate by liquid chromatography tandem mass spectrometry. Chem Res Toxicol 26:926–936

Gamache PH, Meyer DF, Granger MC, Acworth IN (2004) Metabolomic applications of electrochemistry/mass spectrometry. J Am Soc Mass Spectrom 151:717–1726

Garg D, Gandhi T, Gopi Mohan C (2008) Exploring QSTR and toxicophore of hERG K^+ channel blockers using GFA and HypoGen techniques. J Mol Graph Model 26:966–976

Garg M, Chauhan M, Singh PK, Alex JM, Kumar R (2015) Pyrazoloquinazolines: synthetic strategies and bioactivities. Eur J Med Chem 97:444–461

Goldsworthy TL, Reico L, Brown K, Donehower LA, Mirsalis JC, Tennant RW (1994) Transgenic animals in toxicology. Toxicol Sci 22:8–19

Gopi Mohan C, Gandhi T, Garg D, Shinde R (2007) Computer-assisted methods in chemical toxicity prediction. Mini Rev Med Chem 7:499–507

Graham EE, Walsh RJ, Hirst CM, Maggs JL, Martin S, Wild MJ, Wilson ID, Harding JR, Kenna J, Peter RM (2008) Identification of the thiophene ring of methapyrilene as a novel bioactivation-dependent hepatic toxicophore. J Pharmacol Exp Ther 326:657–671

Greene N (2002) Computer systems for the prediction of toxicity: an update. Adv Drug Deliv Rev 54:417–431

Hansch C, Fujita T (1964) ρ-σ-π Analysis. A method for the correlation of biological activity and chemical structure. J Am Chem Soc 86:1616–1626

Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Mol Divers 10:147–158

Helma C, Kazius J (2006) Artificial intelligence and data mining for toxicity prediction. Curr Comput Aided Drug Des 2:123–133

Hsu C, Lin C (2002) A comparison of methods for multiclass support vector machines, neural networks. IEEE Trans 13:15–425

Jeong H (1999) Inhibition of cytochrome P450 2E1 expression by oleanolic acid: hepatoprotective effects against carbon tetrachloride-induced hepatic injury. Toxicol Lett 105:215–222

Ji ZL, Wang Y, Yu L, Han LY, Zheng CJ, Chen YZ (2006) In silico search of putative adverse drug reaction related proteins as a potential tool for facilitating drug adverse effect prediction. Toxicol Lett 164:104–112

Judson PN (1994) Rule induction for systems predicting biological activity. J Chem Inf Comput Sci 34:148–153

Kalgutkar A, Dalvie D, Obach R, Smith D (2012) Pathways of reactive metabolite formation with toxicophores/-structural alerts. React Drug Metab, 93–129

Kar S, Roy K (2013) Predictive chemometric modeling and 3D-toxicophore mapping of diverse organic chemicals causing bioluminescent repression of the bacterium genus *Pseudomonas*. Ind Eng Chem Res 52:17648–17657

Kaur G, Cholia RP, Mantha AK, Kumar R (2014) DNA repair and redox activities and inhibitors of apurinic/apyrimidinic endonuclease 1/redox effector factor 1 (APE1/Ref-1): a comparative analysis and their scope and limitations toward anticancer drug development: Miniperspective. J Med Chem 57:10241–10256

Kazius J, McGuire R, Bursi R (2005) Derivation and validation of toxicophores for mutagenicity prediction. J Med Chem 48:312–320

King RD, Muggleton SH, Srinivasan A, Sternberg M (1996) Structure-activity relationships derived by machine learning: the use of atoms and their bond connectivities to predict mutagenicity by inductive logic programming. Proc Natl Acad Sci 93:438–442

King RD, Srinivasan A, Dehaspe L (2001) Warmr: a data mining tool for chemical data. J Comput-Aided Mol Des 15:173–181

Klopman G (1992) MULTICASE 1 A hierarchical computer automated structure evaluation program. Quant Struct-Act Relatsh 11:176–184

Klopman G, Saiakhov R, Rosenkranz HS, Hermens JL (1999) Multiple Computer-Automated structure evaluation program study of aquatic toxicity 1: Guppy. Environ Toxicol Chem 18:2497–2505

Kortagere S, Ekins S, Welsh WJ (2008) Halogenated ligands and their interactions with amino acids: implications for structure–activity and structure–toxicity relationships. J Mol Graph Model 27:170–177

Kuschewski J, Hui S, Zak SH (1993) Application of feedforward neural networks to dynamical system identification and control, control systems technology. IEEE Trans 1:37–49

Lee A (2006) Adverse drug reactions, 2nd edn. Pharmaceutical Press, United Kingdom

Lewis D, Ioannides C, Parke DV (1995) A retrospective evaluation of COMPACT predictions of the outcome of NTP rodent carcinogenicity testing. Environ Health Perspect 103:178

Lewis D, Ioannides C, Parke D (1996) COMPACT and molecular structure in toxicity assessment: a prospective evaluation of 30 chemicals currently being tested for rodent carcinogenicity by the NCI/NTP. Environ Health Perspect 104:1011

Liao Q, Yao J, Yuan S (2007) Prediction of mutagenic toxicity by combination of recursive partitioning and support vector machines. Mol Divers 11:59–72

Lu D, Giles K, Li Z, Rao S, Dolghih E, Gever JR, Geva M, Elepano ML, Oehler A, Bryant C (2013) Biaryl amides and hydrazones as therapeutics for prion disease in transgenic mice. J Pharmacol Exp Ther 347:325–338

Merlot C, Domine D, Cleva C, Church DJ (2003) Chemical substructures in drug discovery. Drug Discov Today 8:594–602

Munns AJ, De Voss JJ, Hooper WD, Dickinson RG, Gillam EM (1997) Bioactivation of phenytoin by human cytochrome P450: characterization of the mechanism and targets of covalent adduct formation. Chem Res Toxicol 10:1049–1058

Nakayama S, Atsumi R, Takakusa H, Kobayashi Y, Kurihara A, Nagai Y, Nakai D, Okazaki O (2009) A zone classification system for risk assessment of idiosyncratic drug toxicity using daily dose and covalent binding. Drug Metab Dispos 37:1970–1977

Niazi A, Jameh-Bozorghi S, Nori-Shargh D (2008) Prediction of toxicity of nitrobenzenes using ab initio and least squares support vector machines. J Hazard Mater 151:603–609

Noorlander C, Zeilmaker M, van Eijkeren J, Bourgeois F, Beffers R, Brandon E, Bessems J (2008) Data collection on kinetic parameters of substances. Arch Toxicol 87:767–769

Parke D, Ioannides C, Lewis D (1990) Safety evaluation of drugs and chemicals by the use of computer optimised molecular parametric analysis of chemical toxicity (COMPACT). Alternatives to laboratory animals: ATLA, New York

Piparo EL, Maunz A, Helma C, Vorgrimmler D, Schilter B (2014) Automated and reproducible read-across like models for predicting carcinogenic potency. Regul Toxicol Pharm 70:370–378

Rana A, Alex JM, Chauhan M, Joshi G, Kumar R (2015) A review on pharmacophoric designs of antiproliferative agents. Med Chem Res 24:903–920

Ray O, Broda K, Russo A (2004) A hybrid abductive inductive proof procedure. Logic J IGPL 12:371–397

Richard AM, Gold LS, Nicklaus MC (2006) Chemical structure indexing of toxicity data on the internet: moving toward a flat world. Curr Opin Drug Discov Dev 9:314

Rufer CE, Glatt H, Kulling SE (2006) Structural elucidation of hydroxylated metabolites of the isoflavan equol by gas chromatography-mass spectrometry and high-performance liquid chromatography-mass spectrometry. Drug Metab Dispos 34:51–60

Sanderson D, Earnshaw C (1991) Computer prediction of possible toxic action from chemical structure; The DEREK system. Hum Exp Toxicol 10:261–273

Schymanski EL, Jeon J, Gulde R, Fenner K, Ruff M, Singer HP, Hollender J (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol 48:2097–2098

Sharma M, Sharma P, Mondal S, Garg V (2011) Toxicophore and pharmacophore dependent toxicity: perspective review. Pharmacol Online 1:219–235

Sherwin CM, Christiansen SB, Duncan IJ, Erhard HW, Lay DC Jr, Mench JA, O'Connor CE, Petherick JC (2003) Guidelines for the ethical use of animals in applied ethology studies. Appl Anim Behav Sci 81:291–305

Smellie A, Teig S, Towbin P (1995) Poling: promoting conformational variation. J Comput Sci 16:171–187

Smithing MP, Darvas F (1992) HazardExpert: an expert system for predicting chemical toxicity. In: ACS symposium series American chemical society

Snyder RD, Pearl GS, Mandakas G, Choy W, Goodsaid F, Rosenblum I (2004) Assessment of the sensitivity of the computational programs DEREK, TOPKAT, and MCASE in the prediction of the genotoxicity of pharmaceutical molecules. Environ Mol Mutagen 43:143–158

Stepan AF, Walker DP, Bauman J, Price DA, Baillie TA, Kalgutkar AS, Aleo MD (2011) 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 24:1345–1410

Williams DP, Park B (2003) Idiosyncratic toxicity: the role of toxicophores and bioactivation. Drug Discov Today 8:1044–1050

Williams DP, Antoine DJ, Butler PJ, Jones R, Randle L, Payne A, Howard M, Gardner I, Blagg J, Park BK (2007) The metabolism and toxicity of furosemide in the Wistar rat and CD-1 mouse: a chemical and biochemical definition of the toxicophore. J Pharmacol Exp Ther 322:1208–1220

Xu L, Ball J, Dixon S, Jurs P (1994) Quantitative structure-activity relationships for toxicity of phenols using regression analysis and computational neural network. Environ Toxicol Chem 13:841–851

Yang S (2010) Pharmacophore modeling and applications in drug discovery: challenges and recent advances. Drug Discov Today 15:444–450

Zhou X, Su F, Lu H, Senechal-Willis P, Tian Y, Johnson R, Meldrum D (2012) An FRET-based ratiometric chemosensor for in vitro cellular fluorescence analyses of pH. Biomaterials 33:171–180