

Prediction of Adverse Drug Reactions Using Decision Tree Modeling

F Hammann¹, H Gutmann², N Vogt¹, C Helma³ and J Drewe¹

Drug safety is of great importance to public health. The detrimental effects of drugs not only limit their application but also cause suffering in individual patients and evoke distrust of pharmacotherapy. For the purpose of identifying drugs that could be suspected of causing adverse reactions, we present a structure–activity relationship analysis of adverse drug reactions (ADRs) in the central nervous system (CNS), liver, and kidney, and also of allergic reactions, for a broad variety of drugs ($n = 507$) from the Swiss drug registry. Using decision tree induction, a machine learning method, we determined the chemical, physical, and structural properties of compounds that predispose them to causing ADRs. The models had high predictive accuracies (78.9–90.2%) for allergic, renal, CNS, and hepatic ADRs. We show the feasibility of predicting complex end-organ effects using simple models that involve no expensive computations and that can be used (i) in the selection of the compound during the drug discovery stage, (ii) to understand how drugs interact with the target organ systems, and (iii) for generating alerts in postmarketing drug surveillance and pharmacovigilance.

There is no drug whose effects on a patient are solely beneficial. Adverse drug reactions (ADRs) are undesirable effects that occur even when a drug is administered at the proper dose in the correct manner for an appropriate indication, i.e., even when the drug is used correctly and in good faith. ADRs are part of the larger group of adverse drug events. The latter encompass any noxious effects related to drug therapy. Adverse drug events therefore also include drug overdose, drug–drug interactions, and medication and prescription errors. Strategies to reduce adverse drug events due to (human) error can be more easily devised than strategies to reduce ADRs. For the latter, the burden lies mostly with the pharmaceutical industry to develop safe compounds and test them thoroughly. Pharmacovigilance increases the knowledge of ADRs in daily practice, especially with regard to less frequent adverse events. This requires that a larger population must receive the drug before the effects can be observed. Although ADRs may not be preventable, they can be anticipated.

The avoidance and proper management of ADRs are of great importance to public health, and the public at large shows high interest in drug safety. Unrecognized or underreported ADRs not only cause preventable human suffering and costs to the health-care system but can also unnecessarily undermine the public's faith in (and compliance with) drug therapy. This issue

was addressed by the 110th United States Congress, which passed an act in 2007 to increase postmarketing surveillance of regulated drugs. The US Food and Drug Administration has since started the implementation of the so-called Sentinel Initiative, which integrates data from health-care providers, postmarket surveillance agencies, and other sources.

The technical prerequisites for ADR monitoring and the timely generation of alerts are essentially those of data mining—finding and extracting patterns from potentially huge sources of data using mathematical and statistical methods. It is important to have analytical tools that are simple and robust yet specific. For drug safety assessments, an optimal system is one that would receive structural information about the drug and produce notifications that either rule out any noxious potential or point specifically to ADRs that could be expected to arise. Ideally, such a system would arrive at its notifications in a straightforward way, rather than break the problem down into different smaller tests (e.g., detailed physiological modeling of all processes from drug ingestion to the target structures).

There are many *in vitro* and animal models that screen for toxicological effects, but these often do not translate well into clinical practice. Although *in vitro* assays allow optimization of activities and ligands, the environments within which they are

¹Department of Gastroenterology and Hepatology, University Hospital Basel, University of Basel, Basel, Switzerland; ²Drug Metabolism and Pharmacokinetics, Novartis Institutes for BioMedical Research, Basel, Switzerland; ³Freiburg Center for Data Analysis and Modelling, Albert-Ludwigs-University Freiburg, Freiburg, Germany. Correspondence: J Drewe (juergen.drewe@unibas.ch)

Received 1 September 2009; accepted 27 October 2009; advance online publication 10 March 2010. doi:10.1038/clpt.2009.248

performed rarely conform to conditions in human organisms. Also, solubility and permeability can differ greatly between animals and humans because of the differences in acidity and transit time in animal gastrointestinal tracts.¹ Finally, ADRs that develop only with chronic use of a drug or that impair specific human functions (such as higher cognitive functions) are not readily evaluable by these models. For these reasons, human *in vivo* data are preferable.

Such extensive safety data are not easily obtained. A great and largely untapped source, however, lies in the information collected by regulatory offices and pharmacovigilance sites in the form of ADR reports. For this study, we aimed to create a comprehensive survey of ADR reports for a broad variety of drugs in clinical use and develop computational models for understanding and predicting such reactions. The basis for predictions are numerical features that are easily calculated from nothing but the chemical structure of a drug.

A drug must first reach an organ before it can exert an effect, beneficial or otherwise. With regard to central nervous system (CNS) effects, for instance, a common paradigm holds that drugs must be able to cross the blood–brain barrier (BBB). This is certainly true for substances that act directly on target structures in the CNS, but ADRs such as encephalopathy or delirious states can just as well be caused indirectly, for example, by hepatotoxic or nephrotoxic compounds that induce liver or kidney failure, respectively. Most neurological ADRs can be strictly assigned to the CNS. With psychiatric reactions, however, this is not as easy. Dependency and addiction, for example, can be caused by the modulation of central receptors (in the case of opioid addiction). However, they may also be of a psychological nature or caused by peripheral effects (e.g., dependency on vasoconstrictor nasal sprays). Therefore, for this study, we chose to label an ADR as CNS related only if its origin is at least predominantly within the CNS.

Drug-induced liver injury is another major complicating factor in drug therapy and an important reason for drug withdrawals from the market, especially in the case of idiosyncratic reactions.² The risk factors for hepatotoxic ADRs (diabetes, obesity, and old age)³ are already common and will gain in importance as the average life span continues to increase. Compromised liver function results in decreased clearance of hepatically metabolized substances and toxic compounds, which eventually accumulate in the liver. Ultimately, hepatic encephalopathy may develop, owing to the accumulation of ammonia and mercaptans. In this manner, hepatotoxic ADRs can produce CNS effects by proxy and should therefore be taken into consideration in the evaluation of a drug's safety for the CNS.

The same holds true for renal ADRs. The kidney is central to the elimination of water-soluble metabolites and xenobiotics. Renal failure can lead to accumulation of primarily nitrogenous compounds (uremia), with devastating effects on the CNS. Nephrotoxic ADRs are seen in relation to cardiovascular drugs (especially angiotensin-converting enzyme inhibitors), aminoglycosides and nonsteroidal anti-inflammatory drugs (direct tubular damage and interstitial nephritis), and a broad range of other drugs.

Lastly, we consider allergic ADRs. These form a group of commonly reported problems, ranging from laboratory findings (e.g., eosinophilia) to local irritations or skin manifestations to very serious complications such as the first manifestations or exacerbation of allergic asthma. Allergic ADRs are therefore also dependent on a patient's medical history and individual susceptibility. Events attributable to medication errors (especially those that were documented only in relation to overdose) were not included.

Quantitative structure–activity relationship studies are based on the similar property principle, which states that similar compounds have similar properties and activities.⁴ The aim is to determine the important properties and to build models that correlate them with an activity. Most methods are numerical in nature and therefore cannot work with molecules depicted as structural graphs. It is therefore necessary to calculate numerical features from the structures. A wide variety of such numerical molecular descriptors exist, ranging from simple qualities such as molecular weight or lipophilicity to counts of structural motifs (e.g., number of aromatic rings or specific substructures such as carboxyl groups) and complex measures of molecular bulk and connectivity (e.g., the Wiener index⁵). Molecules and their descriptors can be analyzed using the techniques of multivariate statistics or those of machine learning. This yields models that can predict the activity of other, not yet investigated (the so-called unseen) molecules. It is important to note that such models can be built without express prior knowledge of the molecular biology at work (e.g., target binding sites, metabolism) because all that is needed is a list of compounds with measurements of their activities.

For this study, we restricted ourselves to decision tree inference (DTI) analysis. DTI models are easily accessible to human interpretation because they consist of a series of decisions based on molecular descriptors to arrive at their prediction. Also, they are easily applied as a virtual screening tool of chemical databases because they can be directly translated into database queries.

RESULTS

Using the Tanimoto coefficient, we calculated the dissimilarity for all compounds in the data set and arrived at a value of 0.85. This rather high value is evidence of the great structural diversity of the drugs in our data set.

Overall, we found very few substances that were not associated with any ADRs of the CNS. However, when ADRs have a psychiatric or psychosomatic character, they become very subjective and more difficult for nonspecialists to evaluate. Furthermore, a great deal of the ADR data that are acquired reflects adverse events during phase I–III clinical trials. Under these artificial conditions, volunteers and patients are confined to trial units and/or constantly monitored. This induces significant stress and often finds expression in “soft” ADRs. Although these may very well be attributable to the study drug, the relationship is seldom clear.

The safest substances in our set, judging by the total number of ADRs, were the vasodilator and anti-alopecia agent minoxidil

(cardiovascular and pulmonary ADRs are known) and lactulose (gastrointestinal but not hepatotoxic ADRs). The drugs with the highest potential to cause ADRs were the HIV protease inhibitor ritonavir, the dibenzazepine carbamazepine, and the calcineurin inhibitor tacrolimus.

Performance for different end points

A summary of corrected classification rates (CCRs) for every end point and DTI algorithm is given in [Table 1](#).

CNS ADRs. A decision tree generated with the chi-squared automatic interaction detector (CHAID) algorithm performed very well at 89.7% accuracy ([Figure 1](#)). Specifically, compounds with fewer than 10 chain atoms, no carboxyl groups, one or no amine groups, absence of azoles, a partial polar surface area (PPSA-1) $<293.7 \text{ \AA}^2$, and $\log P < 3.1$ show the lowest number of CNS ADRs. Except for the low lipophilicity requirement, these are all characteristics of drugs with good BBB permeability. This implies that the safety of these compounds is not due entirely

to the fact that they do not primarily cross the BBB. Another possible explanation for their safety is that the compounds are extruded by efflux mechanisms. Although the activation of efflux pumps can give rise to drug–drug interactions (and, as a consequence, be the source of medication errors and adverse drug events), drug efflux effectively prevents brain penetration and potentially toxic concentrations. One of the major efflux pumps of the BBB, P-glycoprotein (P-gp, MDR1, ABCB1), was the subject of a recent study conducted in our laboratory.⁶ We therefore cross-checked the compounds in both databases and found 49 matches. Only 12 of 25 compounds associated with CNS ADRs (48%), and 4 of 8 compounds without these ADRs (50%), are reported to be P-gp substrates, which indicates that clearance from the CNS is not a sufficient explanation for the safety of the drug. Rather, it seems that the model reflects the important goals of drug discovery, namely, good permeability and a low number of ADRs. This view is further supported by the requirement for moderate to low lipophilicity, which places a limit on accumulation in the fatty matrix and membranes of the CNS. The CNS concentrations of such compounds are easier to manage and predict, especially during chronic use of the drug. Minoxidil, one of the few compounds in our data set that was not associated with any CNS ADRs, is a good example in this regard.

Hepatotoxic ADRs. The best performance was achieved with a classification and regression tree (CART) decision tree model (CCR: 90.2%, [Figure 2](#)). The model is quite complex, spanning 39 nodes and with a total depth of 5 nodes. Most prominently featured are descriptors of charge and lipophilicity.

Table 1 Corrected classification rates (multiplied by 100 to convert to percentages) for the different classes of adverse drug reactions by decision tree inference algorithm used

	CNS (%)	Liver (%)	Kidney (%)	Allergies (%)
CHAID	89.74	87.15	84.71	78.46
CART	88.01	90.22	88.69	78.94

The best values are shown in bold.

CART, classification and regression tree; CHAID, chi-squared automatic interaction detector; CNS, central nervous system.

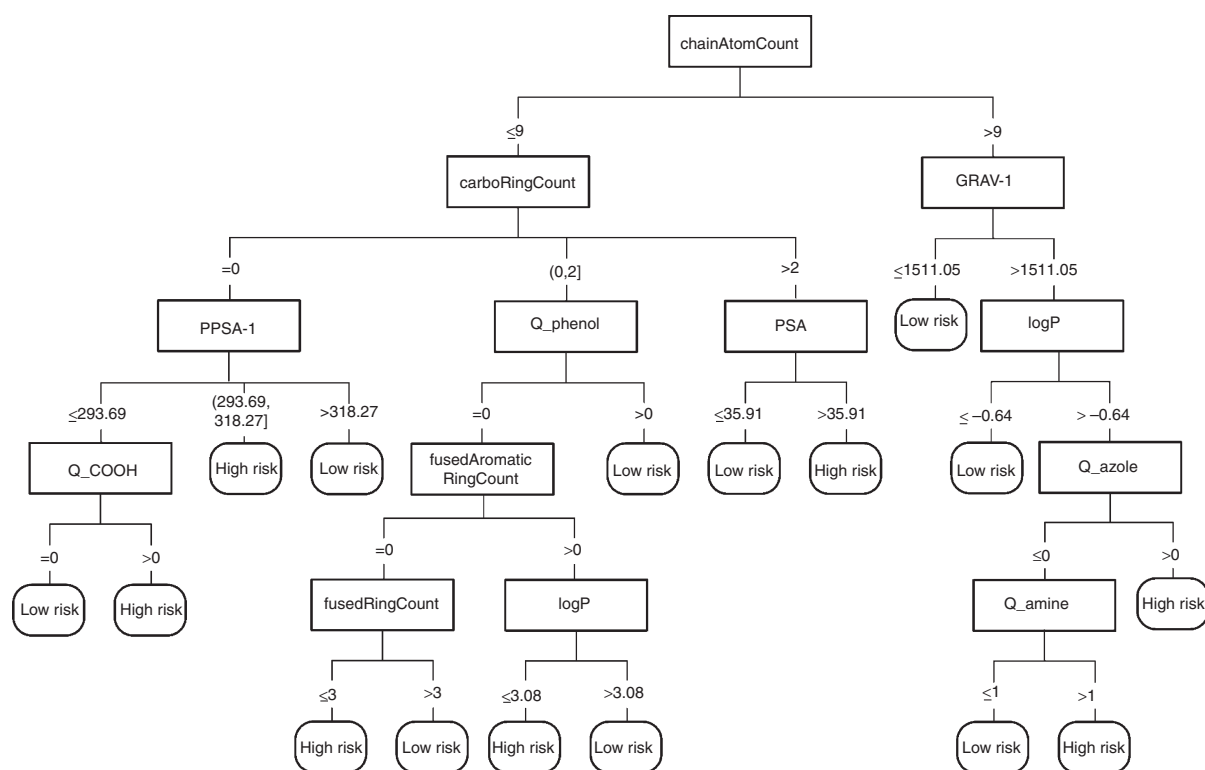


Figure 1 CHAID (chi-squared automatic interaction detector) decision tree model for the central nervous system class of adverse drug reactions for 97 active and 67 inactive compounds ($n = 164$), with a corrected classification rate of 89.74%.

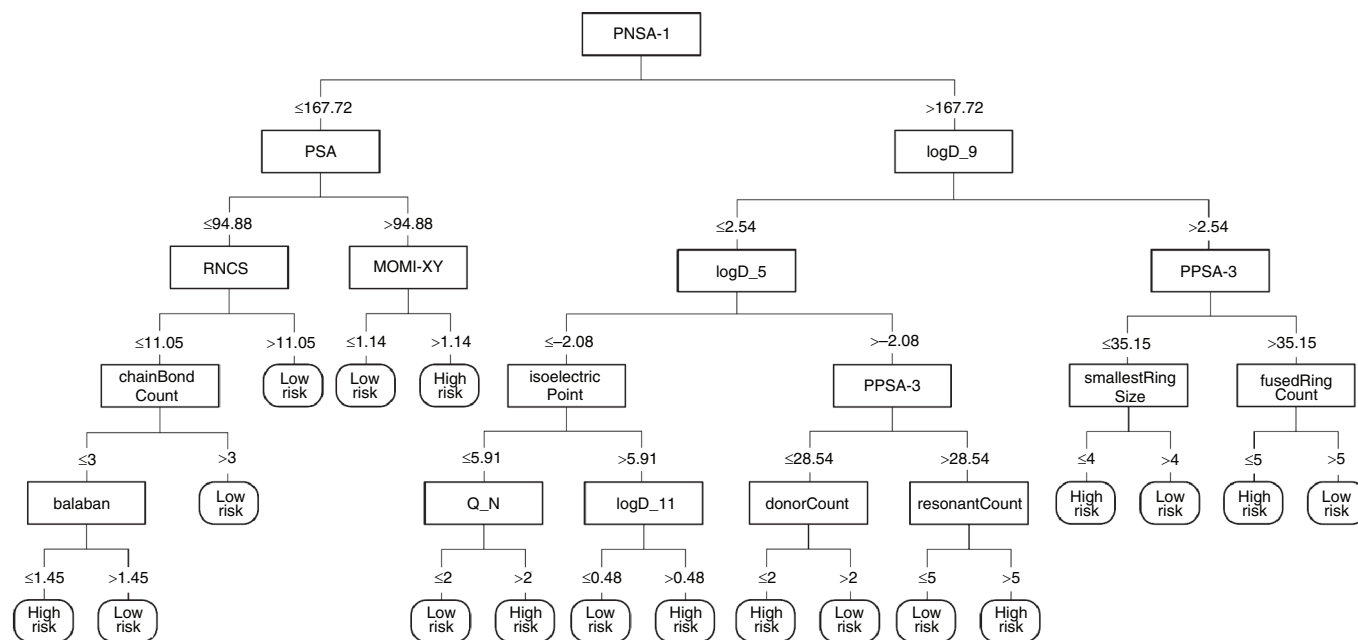


Figure 2 CART (classification and regression tree) decision tree model for the hepatic class of adverse drug reactions for 109 active and 177 inactive compounds ($n = 286$), with a corrected classification rate of 90.22%.

The important characteristics of safe compounds were: a small partial negative surface area ($\text{PNSA-1} < 167.72 \text{ \AA}^2$), small atom-weighted partial positive surface area ($\text{PPSA-3} < 35.16$), small polar surface area ($\text{PSA} < 94.88 \text{ \AA}^2$), and low nitrogen count (< 3). Interestingly, the counts of important substructures such as azoles and nitro groups were not deemed relevant in this model. The CHAID decision tree model (CCR: 87.15%) showed similar descriptors; in addition, it found low counts of sulfur atoms and carboxyl and amine groups, as well as a hydrogen bond donor site count of less than four. The CHAID model places importance on low lipophilic partitioning. Small molecules such as pyridoxine (a form of vitamin B₆ given to lessen the toxic effects of the isoniazid) and the histamine H₁ receptor antagonist meclizine (used in the treatment of motion sickness, vertigo, and nausea during pregnancy) are examples of substances with little potential for causing hepatic ADRs.

This is in line with the common understanding that hepatotoxicity is produced mainly by lipophilic compounds, which not only have higher permeability but also accumulate more easily. The substructure motifs identified by the CHAID algorithm are the basis of toxic metabolites. It seems wise to avoid compounds that can be subject to metabolic activation and can therefore become a source of noxious electrophilic species.⁷ Ritonavir, for example, is rich in nitrogen and has large PSA and PNSA-1 values, as does atazanavir. Both are lipophilic and are noted for their detrimental effects on the liver.

Nephrotoxic ADRs. The CHAID and CART (Figure 3) decision tree models performed well (CCRs: 84.7 and 88.6%, respectively). Again, low PSA ($< 93.78 \text{ \AA}^2$) is an important factor for safety. Few aromatic atoms (< 19), a basic $\text{pK}_a < 10.71$, van der Waals surface area $< 1,014.5 \text{ \AA}^2$, and logP values > 2.43 were seen in the CHAID model for safer substances. The CART model

additionally points out the detrimental influence of amine functions, sulfur, and carboaromatic ring structures.

Despite the comparatively high threshold for van der Waals surface area, the models associate lipophilic and nonpolar compounds with good renal safety. Such compounds are more likely to undergo hepatic rather than renal elimination; as a consequence, they will not have great exposure to renal parenchyma. Tacrolimus appeared to be one of the least safe compounds with regard to kidney- and liver-related ADRs. This was not surprising,⁸ and the compound has many of the structural features that our models look for (high van der Waals surface area, nitrogen count, and logP).

Allergy ADRs. As compared to the other end points, CCRs for this class were lower. Satisfactory models at almost equal CCRs were provided by the decision tree algorithms (CHAID: 78.4%, CART: 78.9%). A diagram of the CART model is given in Figure 4. Both these models also place emphasis on topological complexity, unlike the mostly permeability- and substructure-oriented models described above. Of note are descriptors such as the third Kier κ shape index (which relates structural complexity to extreme shapes of chemical graph theory), the Wiener index (a measure of the degree of branching), and relative moments of inertia (which reflect the degree of symmetry; e.g., MOMIXY and MOMIXZ).

Still, the predictive accuracy is somewhat impressive, given the great structural diversity of compounds with antigenic potential. It is likely that these models owe their performance precisely to this circumstance. Instead of basing their predictions on known antigenic motifs (which were largely not determined by the descriptor set used), both models rely on abstract measures of complexity and show that simple, less branched compounds are safer. This is in line with the common immunological understanding that complex proteins are far more immunogenic than

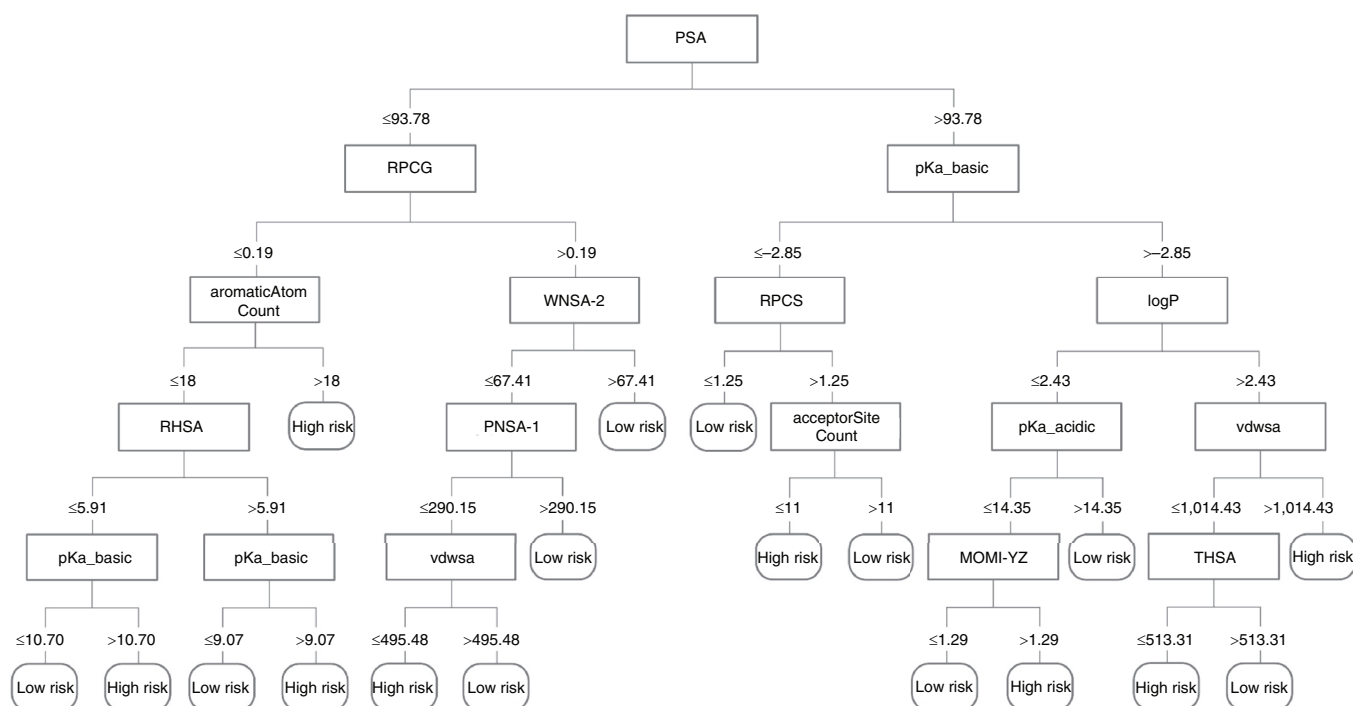


Figure 3 CART (classification and regression tree) decision tree model for the renal class of adverse drug reactions for 126 active and 208 inactive compounds ($n = 334$), with a corrected classification rate of 88.69%.

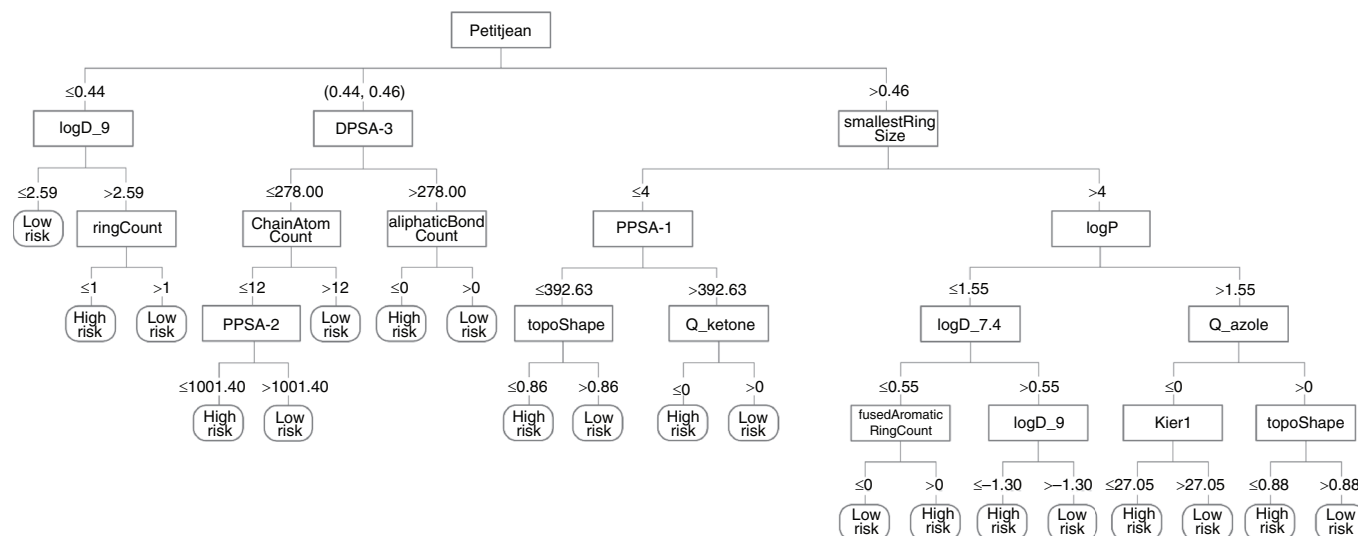


Figure 4 CART (classification and regression tree) decision tree model for the allergy class of adverse drug reactions for 99 active and 239 inactive compounds ($n = 338$), with a corrected classification rate of 78.94%.

polysaccharides. Also, smaller and less complex molecules do not allow for the repetitiveness or structural heterogeneity that triggers an immunological response. The models are therefore in agreement with the observation that very sophisticated drug molecules such as monoclonal antibodies are potential sources of allergic ADRs.

DISCUSSION

In this study, we performed quantitative structure–activity relationship analyses using DTI for compounds that cause

several classes of ADRs. The occurrence of an ADR may also be considered a surrogate marker for a compound's ability to interact with the target organ *in vivo* in humans (data that are notoriously difficult to acquire), and each of the models uses very different features to arrive at decisions. This suggests that they do not model merely the general characteristics of drug likeness but specifically their end points. Our models can therefore be directly used as screening tools in late drug discovery as structural guidelines, in rational drug design, and in pharmacovigilance evaluations. However, it should be noted

Table 2 Summary of the number of compounds per adverse drug reaction (ADR) class and by activity

Class	Lower threshold	Upper threshold	High-risk compounds	Low-risk compounds	Compounds of undetermined activity
CNS	1	20	97	67	343
Hepatotoxicity	0	4	109	177	221
Nephrotoxicity	0	3	126	208	173
Allergies	0	2	99	239	169

CNS, central nervous system.

The number of reported ADRs for high-risk compounds is equal to or greater than the upper threshold, whereas the number of ADRs reported for low-risk compounds is lower than or equal to the lower threshold.

Table 3 Overview of descriptors ($n = 118$) used in the study, by origin (ChemAxon: $n = 58$, CDK: $n = 61$) and class

Class	ChemAxon	Chemistry development kit
Charge analysis	Hydrogen bond acceptor and donor counts and sites (acceptorCounts, acceptorSiteCount, donorCount, donorSiteCount), DREIDING energy, partitioning coefficients for pH 0–14 (logD_0, logD_1, logD_2, logD_3, logD_4, logD_5, logD_6, logD_7, logD_74, logD_8, logD_9, logD_10, logD_11, logD_12, logD_13, logD_14), molecular polarizability (molPol), topological surface area (topologicalSurfaceArea), van der Waals surface area (vdwsa)	Charged partial surface descriptors (PPSA-1, PPSA-2, PPSA-3, PNSA-1, PNSA-2, PNSA-3, DPSA-1, DPSA-2, DPSA-3, FPSA-1, FNSA-1, FPSA-2, FNSA-2, FPSA-3, FNSA-3, WPSA-1, WNSA-1, WPSA-2, WNSA-2, WPSA-3, WNSA-3, RPCG, RNCG, RPCS, RNCs), ²² partitioning coefficient (XlogP)
Constitutional	Counts of atoms, rings, and bonds (aliphaticAtomCount, aliphaticBondCount, aromaticAtomCount, aromaticBondCount, aromaticRingCount, asymmetricAtomCount, atomCount, bondCount, carboaromaticRingCount, carboRingCount, chainAtomCount, chainBondCount, chiralCenterCount, fusedAliphaticRingCount, fusedAromaticRingCount, heteroaromaticRingCount, heteroRingCount, largestRingSize, ringAtomCount, ringBondCount, ringCount, rotatableBondCount, smallestRingSize), molecular refractivity (refractivity), molecular weight (molWeight), resonant count (resonantCount)	Gravitational indexes (GRAV-1, GRAV-2, GRAV-3, GRAV-4, GRAV-5, GRAV-6, GRAVH-1, GRAVH-2, GRAVH-3), ²³ Moment of inertia along the principal axes X, Y, and Z, along with ratios and radius of gyration (MOMIX, MOMIY, MOMIZ, MOMIXY, MOMIXZ, MOMIYZ, MOMIR)
Topological	Balaban index (balabanIndex), weighted Burden matrix (BCUTw1l, BCUTw1h, BCUTc1l, BCUTc1h, BCUTp1l, BCUTp1h), ²⁴ Harary index (hararyIndex), hyper Wiener index (hyperWienerIndex), Platt index (plattIndex), Randic index (randicIndex), Szeged index (szegedIndex)	Kier–Hall κ shape indexes (Kier1, Kier2, Kier3), Petitjean number (PetitjeanNumber) and Petitjean indexes (topoShape, geomShape), Wiener path number and polarity (WPATH, WPOL), Zagreb index (Zagreb)
Motif counts	Q_N (nitrogen), Q_O (oxygen), Q_C (carbon), Q_S (sulfur), Q_benzene (benzene rings), Q_COOH (carboxyl groups), Q_methyl (methyl functions), Q_amine (amines), Q_ketone (ketones), Q_ester (esters), Q_proton_donor (proton donors), Q_phenol (phenol functions), Q_azole (azole functions), Q_nitro (nitro functions), Q_halogen (F, Cl, Br, I)	

that these models are valid only for drug-like molecules and are not necessarily suitable for other areas of application, such as environmental toxicology.

ADRs of the CNS were reported for almost every drug in our list. This is not surprising, given that the assessment of CNS symptoms is often subjective and depends on a patient's self-perception and a clinician's evaluation of it. These ADRs are also influenced to a much greater extent by external factors than are the other ADR classes examined here. Despite this, we were able to provide a thorough estimation tool for this major complicating factor in drug therapy. Models for hepatic ADRs rely on lipophilicity and on counting the substructures. Considering the metabolic role of the liver in bioactivation by unmasking, modifying, and conjugating certain substructures, this is not surprising. Given the complexities of hepatic metabolism, it may be difficult to significantly improve on the performance of general hepatic ADR models such as the ones we present. The models for renal ADRs suggest that safer compounds are less permeable and might not reach sufficiently high

concentrations in the renal parenchyma. Of course, many of the compounds are at least partially eliminated by the kidney, but often substances need to undergo prior metabolic alteration. Better accuracy would therefore be achieved by the model if this factor is also taken into account. However, ADR data such as those used in this study do not permit such an analysis. Finally, the models we present for allergic ADRs are very consistent (although their performance was merely satisfactory). These models rely on abstract measures of the complexity of the compounds, and the results are in line with the current understanding of antigenicity.

We would like to point out that, although our findings are applicable in practice, the requirements presented are more of a disjoint set of rules than a final theory of ADRs. The specific features discussed are used within the models to arrive at a decision; it is not necessary that all features be satisfied for a given compound, and one feature will not necessarily suffice. Furthermore, we have presented models for rather broad classes of ADRs instead of more specific effects. This pooling

was necessary in order to obtain robust and expressive models based on a sufficient number of compounds, but it also reflects the actual pattern of appearance of ADRs in clinical practice. Rarely does one encounter a single detrimental effect in one organ system alone; multiple related effects are the rule rather than the exception.

The different end points presented here are traditionally seen as the sum of different events, namely, transport, metabolism, and docking events. For example, for direct CNS toxicity to occur in a peripherally administered drug, a compound must appear in the peripheral circulation, pass through any of the blood–CNS barriers (most notably, the BBB), face potential active efflux, and be able to interact with a central binding site. Although it may seem appealing from a mechanistic viewpoint to trace this path in terms of a series of physiological models for these events, it must be noted that each prediction is made with a varying degree of uncertainty (which cumulates when an increasing number of models is put in sequence). The analyses in this article, however, prove that it is in fact possible to model such very complex phenomena using simple approaches.

In addition, our models require very little computational power and can easily be applied in high-throughput virtual screenings such as are required in postmarket surveillance. A degree of uncertainty is always associated with the use of predictive models such as the ones presented here. Although our models are built on data related to the effects of the drugs in humans (rather than *in vitro* assays of very specific toxicological mechanisms), they are applicable only to small-molecule drugs such as those encountered in clinical practice. Our models are meant to guide the decision-making process of professionals in the pharmaceutical industry, clinicians involved in pharmacological therapy, and other experts.

Instead of developing models for individual ADRs, we chose to combine related ADRs to form the larger classes described in this article. It would certainly be desirable to have predictive tools for individual ADRs, especially for serious or potentially fatal entities. However, ADRs ($n = 789$) outnumber available compounds ($n = 507$) by a considerable margin, and modeling of every entity is therefore impractical. Also, specific toxicological effects may be better modeled by analysis of culprit targets (e.g., interactions with the human ether-a-go-go-related channel to screen for QT prolongation).

Different analytical tools can also be applied to our data set. Linear models (e.g., multiple linear regression), artificial neural networks, support vector machines, or any of the other currently favored tools may yield additional information or improved models. Our aim, however, was to give easily interpretable and practically applicable models (a strong point of DTI) along with a principal proof of feasibility.

METHODS

Data acquisition. We based our analysis on a list of Food and Drug Administration–approved small-molecule drugs ($n = 1,436$, date of access: 1 January 2008) from the University of Alberta's DrugBank database.⁹ Currently, DrugBank contains >1,300 Food and Drug Administration–approved small-molecule drugs, along with their chemical, pharmaceutical, and biochemical details. Using the official

Swiss drug registry, we retrieved a list of every ADR ($n = 776$) known for the subset of 507 compounds in DrugBank that are marketed in Switzerland. Every ADR was labeled with one or more of the following tags: “CNS related,” “hepatotoxicity related,” “nephrotoxicity related,” “allergy related,” and “miscellaneous.” ADRs that could be classified only as “miscellaneous” were not considered for our analysis.

Unspecific ADRs such as fatigue, nausea, headache, and feelings of dizziness were removed from the final data set because these reactions are almost regularly considered in clinical trials to be a stress reaction to an unfamiliar setting and are rarely verified by an external examiner. We also removed data related to instances of dependency, anorexia (as a consequence of nausea), and very rare or insufficiently characterized ADRs such as fever (unless explicitly stated as drug fever; $n = 35$), symptoms of abstinence ($n = 26$), and death ($n = 9$). On the basis of the class labels, we computed the totals of the ADRs for every class and every compound.

For every ADR class, compounds that represented a low risk were defined as those for which the total number of ADRs was equal to or lower than a preset lower threshold level; compounds associated with high risk were defined as those with ADR numbers equal to or greater than a preset higher threshold level. Compounds with ADR counts between the two thresholds were excluded from the analysis for that class. The purpose of using these experimentally determined thresholds was to be able to obtain sharper distinctions between low-risk and high-risk groups of compounds. A summary of thresholds used is given in [Table 2](#). A table of structures and activities in each class and a complete dataset with SMILES and endpoints can be found as [Supplementary Data](#) online.

Assessment of chemical diversity. The structural diversity of a set of compounds can be estimated by plotting substances as points in a high-dimensional space that is spanned by the available descriptors. We used the Tanimoto coefficient¹⁰ as a measure of similarity (see [Supplementary Data](#) online).

DTI. DTI mimics the human learning and classification process by splitting a training data set into progressively smaller subsets, with each level having higher purity with regard to a property of interest than the preceding one.¹¹ The way in which purity is measured is specific to each algorithm. As learning algorithms, we used CART and the CHAID. The CART algorithm¹² optimizes errors on test data during the growth of the decision tree. The major difference in CHAID lies in its use of the χ^2 -test and pre-pruning, that is, in halting the growth of the decision tree during training, before models grow too large.¹³ General growth limits for all models are a maximum depth of 10 levels, with at least 10 instances in the parent nodes and 5 in the child nodes. The Gini coefficient¹⁴ was used as the measure of impurity in the CART algorithm.

Validation of models. To avoid overfitting (i.e., creating overly complex models with very high predictive accuracy on training data, achieved by extracting too many parameters from the known data at the expense of being able to predict unseen test compounds), we used k-fold cross-validation. Usually, the accuracy of predictions is given as the CCR, expressed as the ratio of correct hits to the total number of compounds (see [Supplementary Data](#) online).

Descriptors. Descriptors are numerical attributes computed from a compound's structure. These include elemental analysis (e.g., atom count), charge analysis (e.g., polarizability, ion charge, topological polar surface area), and geometry (e.g., number of aromatic rings, rotatable bonds), as well as partitioning coefficients and miscellaneous other characteristics (e.g., indicators of hydrogen bonding, DREIDING energy¹⁵). We also included connectivity indexes (Randic,¹⁶ Platt,¹⁷ Wiener,⁵ Kier and Hall κ shape,^{18,19} and the Szeged index²⁰), all of which are information-rich measures of molecular connectivity. A comprehensive list of descriptors used in this work is given in [Table 3](#) (for a glossary see [Supplementary Data](#) online).

Software used. ChemAxon Marvin (Marvin 5.0.4, 2008, <http://www.chemaxon.com>) was used for characterizing chemical structures and substructures, and ChemAxon Calculator Plugins were used for calculating structural properties. Additional descriptors were calculated using the open-source cheminformatics package Chemical Development Kit²¹ (version 1.0.4, 2008, <http://sourceforge.net/projects/cdk>). Decision tree models were grown and validated in SPSS 15.0 for Microsoft Windows (SPSS, Chicago, IL).

SUPPLEMENTARY MATERIAL is linked to the online version of the paper at <http://www.nature.com/cpt>

CONFLICT OF INTEREST

The authors declared no conflict of interest.

© 2010 American Society for Clinical Pharmacology and Therapeutics

- Kerns, E.H. & Di, L. *Drug-like Properties: Concepts, Structure Design and Methods: From ADME to Toxicity Optimization* (Academic Press, Amsterdam, The Netherlands, 2008).
- Andrade, R.J., Robles, M., Fernández-Castañer, A., López-Ortega, S., López-Vega, M.C. & Lucena, M.I. Assessment of drug-induced hepatotoxicity in clinical practice: a challenge for gastroenterologists. *World J. Gastroenterol.* **13**, 329–340 (2007).
- Ramachandran, R. & Kakar, S. Histological patterns in drug-induced liver disease. *J. Clin. Pathol.* **62**, 481–492 (2009).
- Johnson, M.A. & Maggiora, G.M. *Concepts and Applications of Molecular Similarity* (Wiley, New York, 1990).
- Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **69**, 17–20 (1947).
- Hammann, F., Gutmann, H., Jecklin, U., Maunz, A., Helma, C. & Drewe, J. Development of decision tree models for substrates, inhibitors, and inducers of p-glycoprotein. *Curr. Drug Metab.* **10**, 339–346 (2009).
- Tang, W. Drug metabolite profiling and elucidation of drug-induced hepatotoxicity. *Expert Opin. Drug Metab. Toxicol.* **3**, 407–420 (2007).
- Fukudo, M. *et al.* Pharmacodynamic analysis of tacrolimus and cyclosporine in living-donor liver transplant patients. *Clin. Pharmacol. Ther.* **78**, 168–181 (2005).
- Wishart, D.S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).
- Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discov. Today* **11**, 1046–1053 (2006).
- Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
- Breiman, L. *Classification and Regression Trees* (Wadsworth, Belmont, CA, 1984).
- Sonquist, J.A. & Morgan, J.N. The detection of interaction effects: a report on a computer program for the selection of optimal combinations of explanatory variables. Monograph no. 35, 296 (Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 1964).
- Gini, C. *Variabilità e mutabilità* (Memorie di metodologica statistica, 1912).
- Mayo, S.L., Olafson, B.D. & Goddard, W.A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **94**, 8897–8909 (1990).
- Randic, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **97**, 6609–6615 (1975).
- Platt, J.R. Influence of neighbor bonds on additive bond properties in paraffins. *J. Chem. Phys.* **15**, 419–420 (1947).
- Kier, L.B. & Hall, L.H. General definition of valence delta-values for molecular connectivity. *J. Pharm. Sci.* **72**, 1170–1173 (1983).
- Kier, L.B. & Hall, L.H. *Molecular Structure Description: The Electrotopological State* (Academic Press, San Diego, CA, 1999).
- Gutman, I., Khadikar, P., Deshpande, N., Kale, P., Dobrynin, A. & Dömötör, G. The Szeged index and an analogy with the Wiener Index. *J. Chem. Inf. Comput. Sci.* **35**, 547–550 (1995).
- Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E. & Willighagen, E. The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **43**, 493–500 (2003).
- Stanton, D. & Jurs, P. Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Anal. Chem.* **62**, 2323–2329 (1990).
- Katritzky, A.R., Lobanov, V.S. & Karelson, M. Correlation of boiling points with molecular structure. 1. A training set of 298 diverse organics and a test set of 9 simple inorganics. *J. Phys. Chem.* **100**, 10400–10407 (1996).
- Pearlman, R.S. & Smith, K.M. Metric validation and the receptor-relevant subspace concept. *J. Chem. Inf. Comput. Sci.* **39**, 28–35 (1999).

Prediction of Adverse Drug Reactions using Decision Tree Modeling

Supplementary Information

Felix Hammann¹, Heike Gutmann², Nadine Vogt¹, Christoph Helma³, Juergen Drewe^{1*}.

¹: Department of Gastroenterology & Hepatology, University Hospital Basel, University of Basel, Switzerland

²: Drug Metabolism and Pharmacokinetics, Novartis Institutes for BioMedical Research, Basel, Switzerland

³: Freiburg Center for Data Analysis and Modelling, Albert-Ludwigs-University Freiburg, Germany

Table SI 1: Glossary of Descriptors

Descriptor	Explanation
acceptorCount	number of hydrogen bond acceptors [1]
acceptorSiteCount	number of hydrogen bond acceptor sites [1]
aliphaticAtomCount	number of aliphatic atoms
aliphaticBondCount	number of aliphatic bonds excluding bonds to hydrogen
aromaticAtomCount	number of aromatic bonds
aromaticBondCount	number of aromatic bonds
aromaticRingCount	number of aromatic rings (determined by the smallest subset of rings, SSSR) [1]
asymmetricAtomCount	number of asymmetric atoms, i.e. atoms with four different ligands
atomCount	total number of atoms
balabanIndex	Balaban distance connectivity index [2], giving a measure of the complexity of bonds within the molecule
BCUT	Burdon-CAS-University of Texas eigenvalues [3], giving a measure of 3D shape and conformation
bondCount	number of bonds within the molecule including those with implicit hydrogens
carboaromaticRingCount	number of carbocyclic rings (aromatic rings with only carbon)
carboRingCount	number of rings containing only carbon
chainAtomCount	number of chain atoms excluding hydrogen
chainBondCount	number of chain bonds excluding hydrogen
chiralCenterCount	number of centers of chirality within the molecule
donorCount	number of hydrogen bond donors
donorSiteCount	number of hydrogen bond donor sites
DPSA-1	difference of PPSA-1 and PNSA-1 [4]
DPSA-2	difference of PPSA-2 and PNSA-2 [4]
DPSA-3	difference of PPSA-3 and PNSA-3 [4]
DREIDING	energy of molecule within the DREIDING force field [5]
FNSA	fractional negative charged surface areas [4]
FPSA	fractional positive charged surface areas [4]
fusedAliphaticRingCount	number of fused aliphatic rings (as determined by the smallest subset of rings, SSSR) [1]
fusedAromaticRingCount	number of fused aromatic rings (as determined by the smallest subset of rings, SSSR) [1]
GRAV	gravitational indices for heavy atoms (GRAV-1, along with its square root (GRAV-2) and cube root (GRAV-3)); for atom pairs, GRAV-4, GRAV-5, and GRAV-6 are calculated analogously [6-8]

GRAVH	gravitational indices for heavy atoms including explicit hydrogens (GRAVH-1, along with its square root (GRAVH-2) and cube root (GRAVH-3)) [6-8]
hararyIndex	Harary topological index [9]
heteroaromaticRingCount	number of aromatic heterocyclic rings (containing at least one non-carbon atom)
heteroRingCount	number of all heterocyclic rings (containing at least one non-carbon atom)
hyperWienerIndex	measure of expansiveness of the molecular graph based on the Wiener index [1]
Kier	Kier-Hall kappa shape indices measuring the connectedness of a molecule over one (Kier1) to three (Kier3) bonds [1, 10]
largestRingSize	number of atoms in the largest ring
logD	adaptation of logP for ionizable compounds giving the degree of partitioning of the molecule between octanol and water at a given pH
logP	logarithm of the partitioning of the molecule within octanol and water
molPol	magnitude of dipole moment induced in the molecule by the application of an external electrical field
molWeight	molecular mass
MOMI	moments of inertia for the x axis (MOMIX), the y axis (MOMIY), the z axis (MOMIZ), for the principal planes (MOMIXY and MOMIXZ), as well as the radius of gyration around the center of gravity (MOMIR)
PetitjeanNumber	measure of the anisometry of a molecule, relating its diameter to its radius [1, 11]
plattIndex	measure of connectivity, calculated by summing up all entries in edge adjacency matrix [1, 12]
PNSA-1	partial negative surface area [4]
PNSA-2	total charge weighted PNSA-1 [4]
PNSA-3	atomic charge weighted PNSA-1 [4]
PPSA-1	partial positive surface area [4]
PPSA-2	total charge weighted PPSA-1 [4]
PPSA-3	atomic charge weighted PPSA-1 [4]
Q_amine	count of amine groups
Q_azole	count of azole groups
Q_benzene	count of benzene groups
Q_COOH	count of carboxy groups
Q_ester	count of ester functions
Q_halogen	count of halogen atoms
Q_ketone	count of ketone functions
Q_methyl	count of methyl groups
Q_N	count of nitrogen atoms

Q_nitro	count of nitro groups
Q_O	count of oxygen atoms
Q_phenol	count of phenol groups
Q_S	count of sulfur atoms
randicIndex	Randic index of connectivity, giving a measure of branching [13]
refractivity	refractivity index giving the relation of molecular weight to density [1]
resonantCount	number of resonant structures
ringAtomCount	number of atoms located in rings
ringBondCount	number of bonds within ring structures
ringCount	number of all rings in molecule
RNCG	relative negative charge, the ratio of the charge of the most negative atom and the sum of negative charges [4]
RNCS	relative negative charge surface area [4]
rotatableBondCount	number of rotatable bonds
RPCG	relative positive charge, the ratio of the charge of the most positive atom and the sum of positive charges [4]
RPCS	relative positive charge surface area [4]
smallestRingSize	number of atoms within the smallest ring, excluding hydrogens
szegedIndex	measure of connectivity derived from the hydrogen depleted graph [14]
topologicalSurfaceArea	polar surface area as calculated from the contribution of individual fragments [15]
van der Waals surface area	van der Waals surface area in Angstrom
WNSA	surface weighted negatively charged partial surface areas [4]
WPATH	Wiener path number, giving a measure of connectivity by calculating the half of all entries in the distance matrix [1, 16]
WPOL	Wiener polarity number, calculated just as WPATH, however only entries equal to 3 are considered [1]
WPSA	surface weighted positively charged partial surface areas [4]
zagreb	Zagreb index of connectivity, based on the vertex degrees [1, 17]

References

1. Todeschini, R. and V. Consonni, eds. *Handbook of Molecular Descriptors*. Methods and Principles in Medicinal Chemistry, ed. R. Mannhold, H. Kubinyi, and H. Timmerman. 2000, Wiley-VCH: Weinheim, New York.
2. Balaban, A.T., *Distance Connectivity Index*. Chem Phys Letters, 1982. **89**: p. 399-404.

3. Burden, F., *Molecular Identification Number for Substructure Searches*. J Chem Inf Comput Sci, 1989. **29**: p. 225-227.
4. Stanton, D. and P. Jurs, *Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies*. Analytical Chemistry, 1990. **62**(21): p. 2323-2329.
5. Mayo, S.L., B.D. Olafson, and W.A. Goddard, *DREIDING: A Generic Force Field for Molecular Simulations*. Journal of Physical Chemistry, 1990. **94**: p. 8897-8909.
6. Katritzky, A.R. and E.V. Gordeeva, *Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research*. J Chem Inf Comput Sci, 1993. **33**(6): p. 835-57.
7. Katritzky, A.R., et al., *Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics*. Journal of Physical Chemistry, 1996. **100**(24): p. 10400-10407.
8. Katritzky, A.R., et al., *Interpretation of quantitative structure-property and -activity relationships*. J. Chem. Inf. Comput. Sci., 2001. **41**(3): p. 679-85.
9. Invancic, O., T.S. Balaban, and A.T. Balaban, *Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices*. J Math Chem, 1993. **12**(1): p. 309-318.
10. Leach, A.R., *Molecular Modelling - Principles and Applications*. 2nd ed. 2001: Prentice Hall.
11. Petitjean, M., *Applications of the Radius-Diameter Diagram to the Classification of Topological and Geometrical Shapes of Chemical Compounds*. J Chem Inf Comput Sci, 1992. **32**: p. 331-337.
12. Platt, J.R., *Influence of Neighbor Bonds on Additive Bond Properties in Paraffins*. Journal of Chemical Physics, 1947. **15**(6): p. 419-420.
13. Randic, M., *Characterization of molecular branching*. Journal of the American Chemical Society, 1975. **97**(23): p. 6609-6615.
14. Khadikar, P.V., et al., *The Szeged Index and an Analogy with the Wiener Index*. J Chem Inf Comput Sci, 1995. **35**(3): p. 547-550.
15. Ertl, P., B. Rohde, and P. Selzer, *Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties*. J Med Chem, 2000. **43**(20): p. 3714-7.
16. Wiener, H., *Structural Determination of Paraffin Boiling Points*. Journal of the American Chemical Society, 1947. **69**(1): p. 17-20.
17. Gutman, I., *Topological Formulas for Free-Valency Index*. Croat. Chem. Acta, 1978. **51**: p. 29-33.