GENOTOXICITY AND CARCINOGENICITY

CrossMark

Computational identification of structural factors affecting the mutagenic potential of aromatic amines: study design and experimental validation

Svetoslav H. Slavov[1] · Iva Stoyanova-Slavova[1] · William Mattes[1] · Richard D. Beger[1] · Beat J. Brüschweiler[2]

Abstract

A grid-based, alignment-independent 3D-SDAR (three-dimensional spectral data–activity relationship) approach based on simulated ^{13}C and ^{15}N NMR chemical shifts augmented with through-space interatomic distances was used to model the mutagenicity of 554 primary and 419 secondary aromatic amines. A robust modeling strategy supported by extensive validation including randomized training/hold-out test set pairs, validation sets, "blind" external test sets as well as experimental validation was applied to avoid over-parameterization and build Organization for Economic Cooperation and Development (OECD 2004) compliant models. Based on an experimental validation set of 23 chemicals tested in a two-strain *Salmonella typhimurium* Ames assay, 3D-SDAR was able to achieve performance comparable to 5-strain (Ames) predictions by Lhasa Limited's Derek and Sarah Nexus for the same set. Furthermore, mapping of the most frequently occurring bins on the primary and secondary aromatic amine structures allowed the identification of molecular features that were associated either positively or negatively with mutagenicity. Prominent structural features found to enhance the mutagenic potential included: nitrobenzene moieties, conjugated π-systems, nitrothiophene groups, and aromatic hydroxylamine moieties. 3D-SDAR was also able to capture "true" negative contributions that are particularly difficult to detect through alternative methods. These include sulphonamide, acetamide, and other functional groups, which not only lack contributions to the overall mutagenic potential, but are known to actively lower it, if present in the chemical structures of what otherwise would be potential mutagens.

Keyword Aromatic amines · Mutagenicity · Ames assay · 3D-SDAR · Molecular modeling

✉ Svetoslav H. Slavov
Svetoslav.Slavov@fda.hhs.gov

1 Division of Systems Biology, National Center for Toxicological Research, US Food and Drug Administration, 3900 NCTR Rd., Jefferson, AR 72079, USA

2 Federal Food Safety and Veterinary Office FSVO/BLV, Risk Assessment Division, Schwarzenburgstrasse 155, Berne 3003, Switzerland

Introduction

Worldwide, about 15% of total human deaths are cancer related—it is the second leading cause of death in developed countries (following the cardiovascular diseases)—and the third leading cause of death in the developing world following cardiovascular, infectious, and parasitic diseases (Stewart and Wild 2014). Cancer is caused by both intrinsic (such as inherited genetic mutations, hormonal imbalance, and immune conditions) and environmental factors (exposure to synthetic and natural chemicals, infectious organisms, or smoking). During the past few decades, significant progress has been made in understanding, limiting the effect of and preventing cancers induced by exposure to carcinogenic chemicals and in particular to genotoxic carcinogens (Benigni and Bossa 2011). The ability of these chemicals to cause permanent damage to the genetic material of cells is

primarily evaluated by in vitro Ames mutagenicity assays (Ames et al. 1975). Since mutagenicity is tied to the multistep process of carcinogenicity and has been demonstrated to correlate well with it (Benigni 1989), the Ames test has become a preferred early indicator of potential genotoxic carcinogenicity (Gadaleta 2016). Although invaluable in evaluating the mutagenic potential of chemicals, the Ames assay requires a reasonable quantity of a compound (~ 0.75–1 g), and as such, it might be problematic for early stage drug development or testing of impurities. Hence, where bacterial mutagenicity experimental data are absent or impractical, accurate computational prediction of the mutagenic potential of chemicals is highly desirable and may speed up toxicity profiling and evaluation of drug impurities as outlined in ICH Guideline M7 (2015).

Due to reliability concerns often associated with the use of single models, the ICH Guideline M7 recommends a more stringent approach based on consensus between two complementary (Quantitative) Structure–Activity Relationship Approaches ((Q)SARs)—one generated by a statistical system and another utilizing expert rules. There are several statistical-based systems in widespread use: Sarah Nexus from Lhasa Limited (Barber et al. 2016), Leadscope Model Applier—Statistical Models (Yang et al. 2008), CASE Ultra/MC4PC from MultiCASE Inc. (Saiakhov and Klopman 2008), Lazar (Helma 2006), and T.E.S.T (Martin 2016). Among the expert rule-based systems, popular are Lhasa Limited's Derek Nexus (Marchant et al. 2008), Leadscope Model Applier–Expert Alerts (Yang et al. 2008) and some of the modules in Toxtree (Patlewicz et al. 2008a).

Due to their use as starting materials in the pharmaceutical industry, which subsequently leads to their presence as impurities in many drug formulations, the primary and secondary aromatic amines (further referred to as PAA and SAA) are prime candidates for mutagenicity evaluation (McCarren and Springer 2011). Such scrutiny is further supported by a substantial amount of evidence linking the aromatic amines and their metabolites to well-known mechanisms of DNA damage such as the formation of DNA adducts with the nitrenium ions, direct DNA intercalation and the formation of reactive oxygen species resulting in oxidative stress (Benigni and Bossa 2011). For these reasons, structural alerts warranting further investigation are often designed around the presence of amino-aromatic moieties. The overly general nature of such alerts, however, has been demonstrated to lead to a significant amount of false positives for certain subclasses of chemicals (Cariello 2002). As an alternative to the use of structural alerts, numerous QSARs utilizing molecular descriptors have been proposed during the past few decades, some of which are summarized in Table 1.

Although some of the published models were properly derived and validated, many suffered from: (1) absence of or poor validation; (2) omission of data critical for their reproducibility; (3) lack of structural/mechanistic interpretation; and (4) undefined applicability domains. Hence, it is not surprising that none of these approaches (except those based on nitrenium ions stabilities) is routinely used in estimating the Ames mutagenicity of aromatic amines. To avoid the above outlined shortcomings and to build an OECD compliant model, this work was focused on:

1. developing a transparent modeling procedure with well-defined training/hold-out, validation, and "blind" external test sets;
2. utilizing a rigorous modeling procedure relying on multiple fully randomized training/hold-out set pairs processed by a robust algorithm designed to avoid overfitting;
3. defining an optimal prediction space/applicability domain based on a similarity score calculated from the 3D-SDAR fingerprints;
4. using 3D-SDAR mapping techniques to facilitate the identification of structural features associated either positively (enhancers) or negatively (reducers) with mutagenicity.

Data sets

Curated data sets consisting of 579 PAA and 437 SAA provided to the National Center for Toxicological Research (NCTR) by Lhasa Limited were preprocessed using the following rules: (1) all solvent molecules were discarded to obtain a single structure; (2) all counter-ions of organic salts such as Cl^-, Br^-, SO_4^{2-}, Na^+, etc., were removed; and (3) in cases of mixtures, only the active component identified by its CAS number was retained. The removal of counter-ions from the PAA set resulted in 15 duplicates and 2 triplicates. In cases of concordant calls, one of the duplicates/triplicates was retained, whereas in cases of conflicting calls, both chemicals were discarded. In total, 25 PAAs were removed from the data set. Similarly, the identification of 15 SAA duplicates resulted in the removal of 18 chemicals. After the "washing" procedure described above, the PAA and SAA data sets were comprised, respectively, of 554 and 419 chemicals; these will be referred to as initial data sets.

To build reliable predictive models, the initial data sets of 554 PAA and 419 SAA were split into modeling (3/4) and validation (1/4) subsets using the following strategy designed to generate validation sets representative of the modeling sets. Since the conjugated systems in the molecules of aromatic amines are essential for mutagenicity, the Toxmatch (Patlewicz et al. 2008b) software was used to calculate "the number of atoms in the largest pi system" descriptor, which ranked values determined the assignment of each chemical to either the modeling or the validation

Table 1 Summary of published models on the mutagenicity of aromatic amines

Endpoint	Data set size	Data processing algorithm	Descriptors	Statistical parameters	Reference
Mutagenic activity	88	MLR	$\log P$, E_{LUMO}, E_{HOMO}, I_1	$R = 0.898$, $s = 0.86$	Debnath (1992)
Mutagenic potency	165	MLR	HOMO, LUMO, number of fused aromatic rings, etc	$R^2_{\text{adj}} = 0.62$, $F = 91.7$	Hatch (1997)
Revertants/nanomole	95	MLR	CODESSA descriptors	$R^2 = 0.834$, $R^2_{\text{cv}} = 0.81$	Maran (1999)
Revertants/nanomole	95	Hierarchical QSAR	Topological, geometric and QM indices	$R = 0.89...0.92$ $F = 37.8...57.2$	Basak (2001)
Revertants/nanomole	95	MLR	Electrotopological state indices	$R = 0.87$, $F = 44.2$	Cash 2001
Revertants/nanomole	Trn = 47/Tst = 48	Least squares	Graphs of atomic orbitals	$R_{\text{trn}} = 0.87$, $R_{\text{tst}} = 0.87$ $s = 1.05$, $s = 0.86$	Toropov (2001)
Revertants/nanomole	Trn = 60/Tst = 39	Ordinary least squares, GA	DRAGON descriptors	$R_{\text{trn}} = 0.80$, $R_{\text{tst}} = 0.69$	Gramatica (2003)
Revertants/nanomole	95	Counter propagation neural network	Topostructural, topochemical, geometrical and quantum chemical	$Q^2_{\text{LOO}} = 0.751$; $F = 279.9$	Vracko (2004)
Revertants/nanomole	95	MLR, ANN	Diatomic fragments	$R^2 = 0.77...0.91$ $Q^2_{\text{LOO}} = 0.67...0.70$	Casalegno (2006)
Revertants/nanomole	20	PLS, MLR	BMC chromatographic retention, molar refractivity, MW, etc	$R^2 = 0.86$, se = 0.8, $F = 52$	Torres-Cartas (2007)
Revertants/nanomole	Trn = 97/Tst = 25	HSVR, SVM, ANN, GFA	Discovery Studio and E-Dragon descriptors	$R^2_{\text{trn}} = 0.78...0.93$ $R^2_{\text{tst}} = 0.73...0.85$	Leong (2010)
Ames active/inactive	Trn = 212 Calibration = 71 Validation = 71	Regression, kNN	SMILES descriptors, similarity indices, CAESAR-VEGA descriptors	ACC = 0.77...0.85 SN = 0.76...0.87 SP = 0.73...0.88	Manganelli (2016)
Ames active/inactive	257	$\Delta\Delta E$ cutoff	Nitrenium ions stabilities ($\Delta\Delta E$)	ACC = 0.84...0.86 SN = 0.88...0.93 SP = 0.62...0.77	Bentzien et al. (2010)
Ames active/inactive	846 unique (32, 326, 189, 459)	Formation energy cutoff	Nitrenium ion formation energy	ACC = 0.55...0.88 SN = 0.73...0.90 SP = 0.33...0.88	McCarren (2011)

set. The chemicals in the PAA and SAA sets were arranged in an ascending order of their calculated descriptor values and were labeled sequentially as "a", "b", "c" and "d". The chemicals labeled by "d" were then moved to form the validation set. Thus, the initial data set of 554 PAA was split into 416 chemicals forming the modeling set and 138 compounds comprising the validation set. Respectively, 315 out of the total of 419 SAAs were assigned to the modeling set, whereas the remaining 104 chemicals formed the validation set. Both modeling sets were further (repeatedly) split (by a bagging-like randomization algorithm) into training (80% of the chemicals, 333 in case of PAA, and 252 in case of SAA) and hold-out test (20% of total, 83 in case of PAA, and 63 in case of SAA) subsets. This procedure is described in detail in the "model building" section.

Once the modeling phase was complete and the models were reported back to Lhasa Limited, NCTR was provided with two sets of PAAs and SAAs to be used as "blind"

external test sets. The Tanimoto similarity score (defined as $T = \frac{A.B}{A^2 + B^2 - A.B}$, where A and B are vector rows of the bin occupancies for compounds A and B) was used to identify and retain only those chemicals from the two external test sets that had close analogs among the PAA and SAA training set compounds. Similarity thresholds equal to the average of the maximum pairwise similarities (T_{max}) calculated for all possible pairs of chemicals in the training sets were used to determine which external test set chemicals belong to the applicability domains of the PAA ($T > 0.294$) and SAA ($T > 0.274$) models. In other words, only those chemicals that belonged to the applicability domains of our models were retained. The performance parameters for these two test sets were used to estimate the true predictive power of our models. Under collaboration with the Swiss Federal Food Safety and Veterinary Office (FSVO), NCTR predicted the mutagenic potential of a data set of 23 aromatic amines,

whose activities were not disclosed to the NCTR before the predictions were made. These were tested in a two-strain Ames assay at Envigo Lab (Rossdorf, Germany) (Brüschweiler and Merlot 2017).

Methodology

Endpoint transformation

The PAA and SAA sets provided by Lhasa Limited contained data expressing the mutagenic potential of chemicals as a binary overall call, with 0 (or negative) encoding for a lack of evidence for mutagenicity and 1 (or positive) encoding for a strong evidence of mutagenic activity (see Table 2 and the SI spreadsheet). The overall call was defined as a surrogate endpoint combining Ames assay data derived from TA98 and TA100 *Salmonella typhimurium* strains (with and without metabolic activation) with literature evidence from other salmonella strains and bacteria (e.g., *Escherichia coli*). In addition, the PAA and SAA data sets included records detailing the outcomes of the Ames test for two of the most sensitive to aromatic amines strains, TA98 and TA100 as well as their metabolically activated counterparts TA98+S9 and TA100+S9 (Fan et al. 1998). These two strains are often considered sufficient to detect either frame-shift (TA98) or base-pair substitution (TA100) mutations (Leong et al. 2010). Although there are different variants of the Ames test, most protocols recommend the use of either four (TA97a, TA98, TA100, and TA102) or five (TA98, TA100, TA1535, TA1537, and TA1538) strains (Winder 2004). Of these, strains TA97a, TA98, TA1537, and TA1538 detect frame-shift mutations, strains TA100 and TA1535 are used to identify base-pair mutagens and the TA102 strain detects oxidants and cross-linking agents not detected by other strains (Winder 2004).

The Ames test outcomes for TA98, TA98 + S9, TA100, and TA100 + S9 were categorized as follows: negative (no evidence of mutagenicity), conflicting (evidence for and against), equivocal (absence of strong evidence for mutagenicity), and positive (strong evidence for mutagenicity).

Table 2 Average *P*-scores and distribution of the positive and negative samples between the test and training sets based on the binary overall call for the PAA and SAA data sets

Chemical class	Set	Positive	Negative	Average *P* score
PAA	Modeling	249	167	2.487
	Validation	78	60	2.376
SAA	Modeling	155	160	2.309
	Validation	62	42	2.538

Due to the inherent limitations of models based on binary endpoints (i.e. their inability to identify subtle but important structural determinants of activity) a coherent approach taking advantage of all available data to construct an intermediary, synthetic endpoint (referred to as *P*-score) was employed. The Ames test outcomes reported for the individual TA98, TA98+S9, TA100, and TA100+S9 assays were assigned categorical variables, such that the negative outcomes were encoded as 1, the conflicting as 2, the equivocal as 3, and the positive as 4. A value of 5 (instead of 4) was considered for the positive outcomes, thus positioning the equivocal outcomes in the middle (halfway between the negative and positive outcomes). This idea was discarded due to the increased discontinuity of the transformed *P*-scores, characterized by more sparsely populated or unoccupied regions (see Fig. 1b). For the purpose of this transformation, the original overall call assignments (0/1) were encoded as 1 and 4, representing, respectively, the negative and positive calls. The transformed intermediary endpoint *P* was calculated using the following formula:

$$P = \frac{\frac{\sum_{i=1}^{n} A_i}{n} + C}{2},$$

in which the first term in the numerator is an average of the Ames assay outcomes for TA98, TA98+S9, TA100, and TA100+S9 and *C* is the overall call. As defined, the contribution of the overall call *C* to the intermediary score *P* is equal to that of the four individual assays. There were several benefits resulting from this decision:

1. It helped resolve seemingly conflicting situations, such as the case in which 8 PAAs and 7 SAAs, each characterized by four negative individual Ames assay outcomes (for TA98, TA98 + S9, TA100, and TA100 + S9) were assigned an overall positive call due to evidence for mutagenicity from other strains.
2. It helped to retain chemicals for which only an overall call was reported (44 PAA and 47 SAA).
3. The resulting intermediary endpoint *P* suffered from fewer (and smaller) gaps in continuity—i.e. improved uniformity of the distribution (see Fig. 1a).

Following from the above definition, the *P*-score will increase gradually to its maximum value of 4 as more of its components increase to their maximum values. Since each component of the *P*-score represents a different mechanism (need for metabolic activation, frame-shift mutation, base-pair substitution mutation, mutations caused by oxidants and cross-linking agents, etc.), its increase would be linked to the ability of each individual compound to damage DNA via one, two, or more alternative mechanisms. For example, a chemical with a *P*-score of 1 would be inactive under all
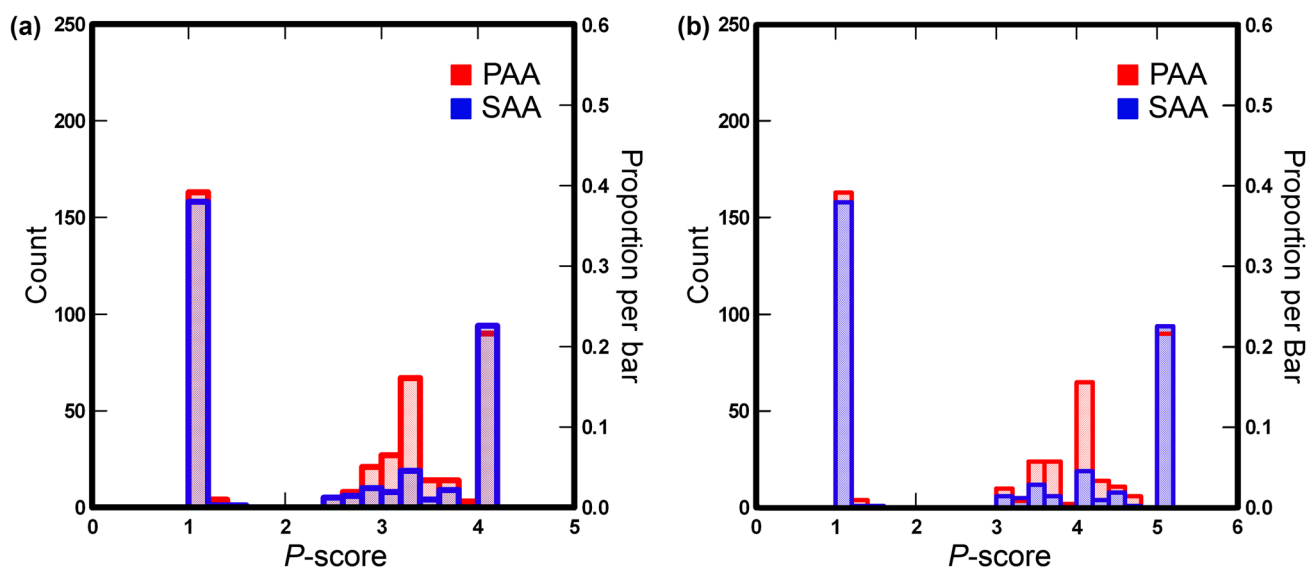
Fig. 1 Histogram of the distribution of the *P*-scores for the PAA and SAA data sets based on the: **a** "1, 2, 3, 4" and **b** "1, 2, 3, 5" categorical class assignments, encoding for the negative, conflicting, equivocal and positive outcomes

testing conditions (different strains and added S9 fraction), whereas a compound that needs a metabolic activation to revert TA98 (but is inactive under other testing conditions) would have a *P*-score of 2.875. On the opposite end of the spectrum would be a compound that is active in both strains (with and without metabolic activation) and for which there might be evidence for mutagenicity from other strains (potential indicators for alternative mechanisms). Such a compound would be characterized by a *P*-score of four (its highest possible value). In general, a higher *P*-score would indicate that a given compound is mutagenic under more testing conditions (for example, active in TA98, TA100 or both with and without metabolic activation and potentially active in other strains) and by extension might damage DNA via multiple alternative mechanisms. Alternative scoring functions that might be more closely related to the number of different mechanisms are also possible; however, as defined, the *P*-score was preferred for its simplicity.

Due to the improved continuity of the *P*-score (as opposed to the binary overall call), the generated 3D-SDAR models should be able to pinpoint specific structural features associated either "positively" or "negatively" with mutagenicity. It is also important to note that at the final stage of the modeling procedure, the predicted *P*-score values were converted back to a binary class assignment (negative/positive) using cut-off values determined from each of the training sets (shown in the last column of Table 2).

3D-SDAR fingerprint generation

3D-SDAR is a grid-based approach, which is based on fingerprints constructed from the NMR chemical shifts of pairs of atoms (determining the *X* and *Y* coordinates of the individual fingerprint elements) and their corresponding through-space atom-to-atom distances (determining the *Z*-coordinate). These fingerprints are further tessellated by regular grids, thus generating descriptor matrices with thousands of columns (see the SI spreadsheet), later processed by a bagging-like Partial Least Squares (PLS) algorithm. The following steps describe the generation of these 3D-SDAR fingerprints:

4. The molecular geometries of all chemicals were optimized using the AM1 Hamiltonian as implemented in Hyperchem v. 8.03 (2007). The ACDLabs NMR (2011) and XNMR (2011) predictors were used to simulate the ^{13}C and ^{15}N NMR chemical shifts;

5. The ranges of the ^{13}C and ^{15}N chemical shifts for the two modeling sets were determined. In case of PAAs, δ^{13}C covered a range between 5.82 and 206.9 ppm, whereas the δ^{13}C of SAAs ranged from 5.07 to 207.99 ppm. The lower bound of the PAAs δ^{15}N range was -373.56 ppm, while the upper bound was 632.13 ppm. Respectively, the lower and upper bounds of SAAs δ^{15}N were -360.69 ppm and 632.13 ppm;

6. To avoid the overlap between the δ^{13}C and δ^{15}N regions, the chemical shifts of the nitrogen atoms were translated downfield by 1000 ppm past the upper bound of the δ^{13}C region. This translation, however, was performed only

for convenience, as to simplify the fingerprint binning procedure and the fingerprint element count algorithm. For the purpose of structural interpretation, the original values of $\delta^{15}N$ were restored once the modeling phase was complete.

7. The through-space distances between all C–C, C–N, and N–N atom pairs were calculated and combined with their corresponding chemical shifts to generate unique fingerprints representing each compound in the 3D-SDAR space. One such fingerprint constructed for the structure of aniline is shown in Fig. 2.

8. The above-described downfield translation of $\delta^{15}N$, allowed division of the *XY*-plane into three distinct regions: carbon–carbon, carbon–nitrogen and nitrogen–nitrogen. Since the gyromagnetic ratio γ of the carbon atoms is approximately 2.5 times that of the nitrogens, the grid density was set in such a way that the bins in the C–N region were 2.5 times larger than the bins in the C–C region and 2.5 times smaller than those in the N–N region. To explore the 3D-SDAR parametric space, tessellations with grid densities ranging from 4 × 4 ppm × 0.5 Å to 10 × 10 ppm × 1 Å in the C–C region, 4 × 10 ppm × 0.5 Å to 10 × 25 ppm × 1 Å in the C–N region and 10 × 10 ppm × 0.5 Å to 25 × 25 ppm × 1 Å in the N–N region (with incremental steps of 0.5 Å on the *Z*-axis and 2 ppm on the chemical shifts plane *XY*) were generated.

9. After tessellation of all individual fingerprints, the fingerprint elements occupying each bin were counted and stored in 3D-SDAR descriptor matrices. In these matrices, each row vector represents a single compound, whereas each column vector contains the counts of the fingerprint elements occupying a specific bin.

Model building

A bagging-like PLS-SIMPLS algorithm (De Jong 1993) written in Matlab (v8.0, 2012) was used to process each of the generated 3D-SDAR descriptor matrices. Before processing, all descriptors were standardized using the "zscore" Matlab function. To avoid overfitting and build models free of selection bias, the PAA and SAA modeling sets (3/4 of the initial sets) were split randomly into training (80% of the chemicals) and hold-out test (the remaining 20% of the chemicals) sets. One hundred such randomizations were performed. On each run, the PLS algorithm fitted the *P*-scores for the training set and predicted the hold-out test and the validation sets (see Fig. 3). At the end, the aggregated predictions for the individual compounds belonging to any of the above defined sets (training, hold-out and validation) were averaged and a threshold equal to the average of the *P*-scores for the training set was used to classify each chemical as either positive or negative. Due to the repeated splitting of the modeling set into training/hold-out test set pairs, each compound from the modeling set had two predicted values; one, when it was randomly assigned to the training set and another, when it belonged to the hold-out test set.
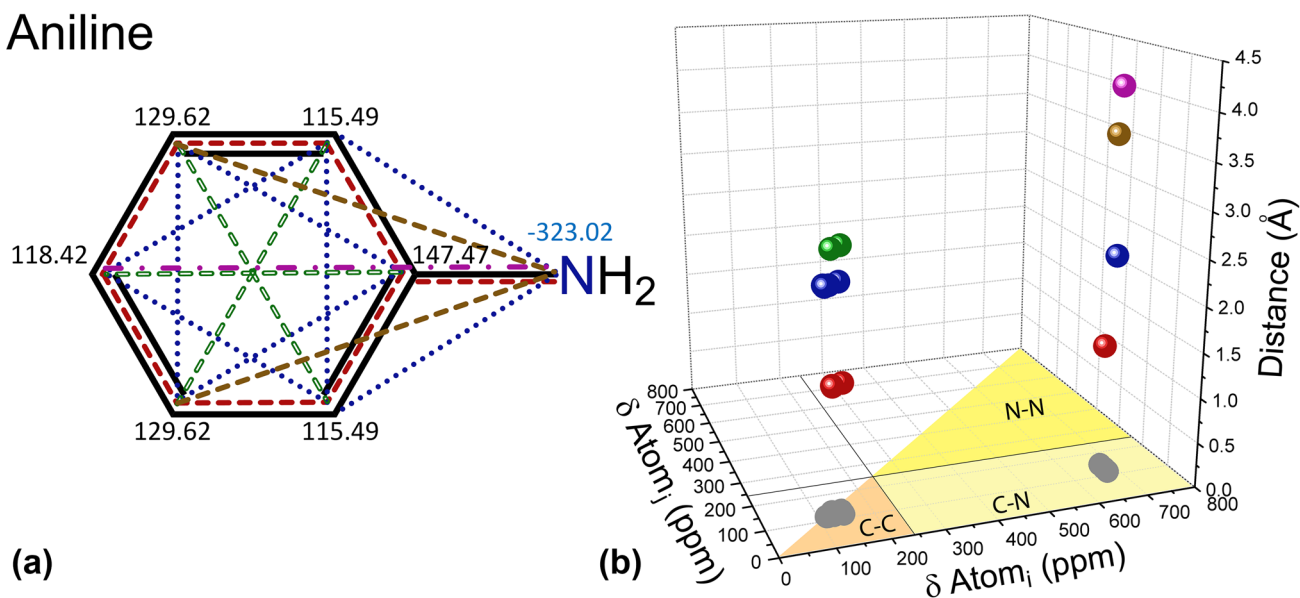

Fig. 2 Molecule of aniline (1a) and its corresponding 3D-SDAR fingerprint (1b). The colors of the dashed lines correspond to the colors of the fingerprint elements shown in 1b. As explained above, $\delta^{15}N$ is translated downfield by 1000 ppm
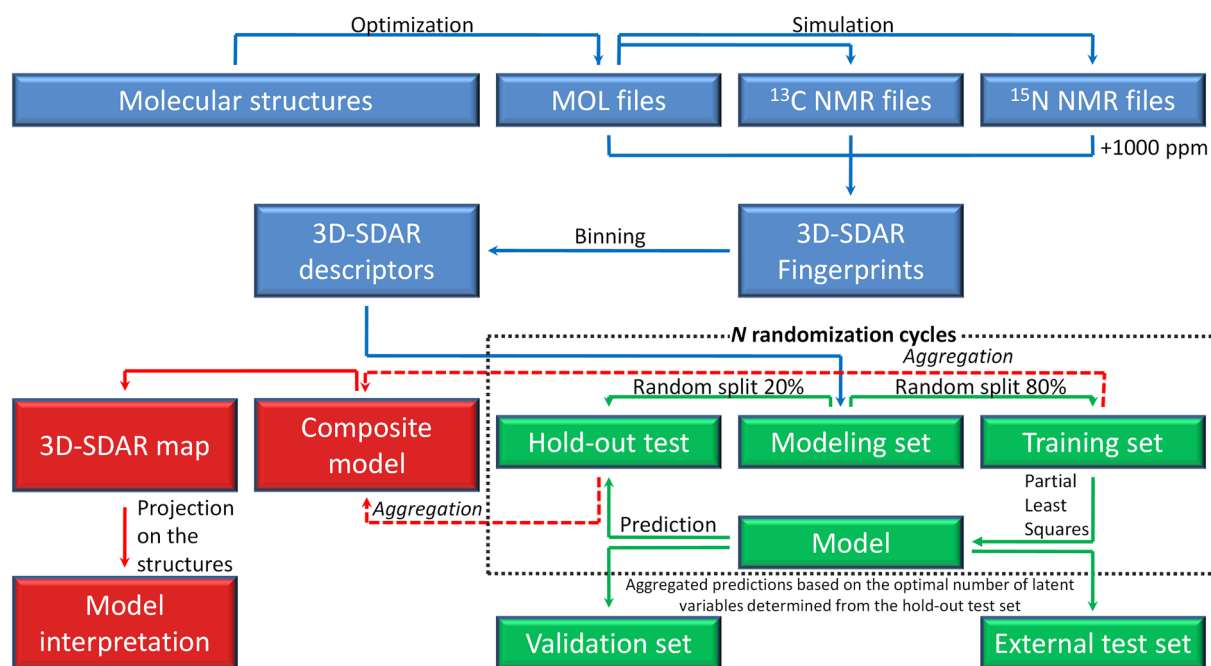
Fig. 3 Flowchart of the 3D-SDAR modeling process

Because the training set predictions merely reflect the ability of PLS to fit data (and are thus irrelevant for estimating the "true" predictive power of models), the Results and discussion section will be focused only on the predictions made for the hold-out, validation and the "blind" external test sets.

Models with between 1 and 10 latent variables (LVs) were generated and the accuracy of prediction for the hold-out test set was used to determine their optimal number. To recreate the same training/hold-out test set sequence when processing 3D-SDAR descriptor matrices based on different tessellations, the random number generator was initialized with a pre-specified random seed. This was also done with the intention of building completely reproducible models.

Interpretation

The linear nature of PLS combined with the robustness of the bagging-like randomization algorithm and the atom-level resolution of 3D-SDAR allow for a straightforward determination of the structural factors playing a role in mutagenicity. Since 3D-SDAR relies on composite models aggregating the predictions from multiple randomized models, the PLS weights from the individual models cannot be utilized directly to rank the 3D-SDAR descriptors by importance. An alternative approach would be to extract a preset number of highly weighted (positive as well as negative) descriptors for a preset number of latent variables on each run and to further calculate the frequency of occurrence of each descriptor as a fraction of the total number of extracted descriptors.

The more frequently a given descriptor appears in the series of models, the more likely it is that this descriptor is truly related to the observed biological effect. During the next stage, the descriptors with the highest frequency of occurrence are mapped on the set of chemical structures and the recurring structural patterns would then represent molecular features that if present would lead to either an increase (positive PLS weights) or a decrease (negative PLS weights) in the overall mutagenic potential.

Results and discussion

Predictive performance

For both PAA and SAA sets, models utilizing grids with 4×4 ppm $\times 0.5$ Å in the C–C region, 4×10 ppm $\times 0.5$ Å in the C–N region and 10×10 ppm $\times 0.5$ Å in the N–N region resulted in models with the highest overall accuracy for the hold-out test sets. In the case of PAA, the highest accuracy of prediction was achieved using models with 7 LVs, whereas in the case of SAA 6 LVs were found to be optimal. Our earlier work (Slavov et al. 2014, 2016; Stoyanova-Slavova et al. 2017), dealing with conversions between continuous and categorical variables indicated that the effect of the continuous variable distribution bias on the class assignment could be mitigated using cut-off values equal to the average of the training set activity vectors (in this case, the average of the *P*-scores for the PAA and SAA training sets). In other

words, to convert the averaged P-scores for each compound back to a binary positive or negative class assignment, the cut-off values shown in the last column of Table 2 were used (2.487 for PAA and 2.309 for SAA). Since the probability to classify correctly chemicals near the above cut-off values decreases as a function of their proximity to the cutoffs, confidence bands excluding all compounds with P-scores within the ± 5% on both sides (2.362...2.611 and 2.194…2.425) of the cut-off values were defined. The performance parameters of the 3D-SDAR models for the hold-out test, validation and external test sets are summarized in Table 3. The statistical parameters for the sets in Table 3 marked with an asterisk exclude the chemicals within the low confidence bands. As can be seen from the performance parameters shown in Table 3 and Fig. 4, both PAA and SAA models offer an excellent performance transferability from the hold-out test set to the validation and the "blind" external test sets.

Although the SAA sets were somewhat smaller in comparison to the PAA sets, the SAA model seemed to perform slightly better when predicting the validation and the "blind" external test set. The SAA model also needed fewer LVs to achieve this somewhat higher performance level. One possible explanation might be in the slightly better (although sparser) coverage of the chemical space by the SAA chemicals, which were less similar to each-other (average $T_{max} = 0.274$) than the chemicals constituting the PAA set (average $T_{max} = 0.294$). Hence, models based on more structurally diverse chemicals might be able to capture trends in the data that are not accessible to models based on sets with more limited structural variability and, respectively, benefit from that information when predicting external test sets. An alternative explanation could be derived from the P-score distribution shown in Fig. 1, which indicates the presence of fewer SAAs with intermediate values of P and m, respectively, fewer "borderline" cases.

Table 3 Predictive performance of the best PAA and SAA models

Performance parameters	PAA (7 latent variables)						SAA (6 latent variables)					
	Hold-out set		Validation set		External test		Hold-out set		Validation set		External test	
N	416	355*	138	109*	121	104*	315	270*	104	86*	91	67*
Accuracy	0.70	0.74	0.79	0.83	0.70	0.72	0.73	0.75	0.84	0.87	0.69	0.76
Sensitivity	0.67	0.72	0.82	0.86	0.64	0.65	0.71	0.72	0.79	0.85	0.68	0.68
Specificity	0.74	0.77	0.75	0.78	0.80	0.84	0.75	0.78	0.90	0.91	0.71	0.88
NPV	0.60	0.66	0.76	0.80	0.57	0.58	0.73	0.74	0.75	0.79	0.57	0.64
PPV	0.79	0.81	0.81	0.85	0.84	0.88	0.73	0.76	0.92	0.94	0.80	0.90

*Dataset size after excluding all chemicals lying within the low confidence bands
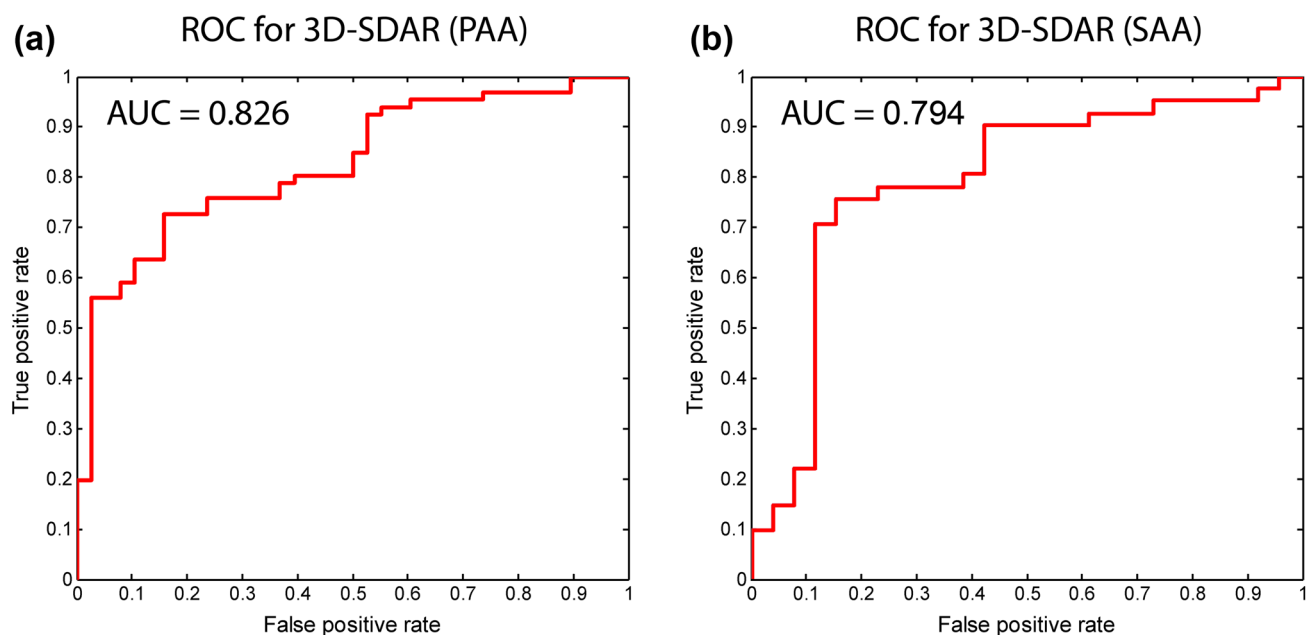

Fig. 4 ROC curves for the primary (**a**) and secondary (**b**) amines prediction sets after removal of the modeling inconclusives near the cutoff

Collaboration between NCTR and FSVO allowed a comparison between the predictions generated by 3D-SDAR, Lhasa Limited's Sarah Nexus and Derek Nexus and a two-strain Ames test performed on a set of 23 aromatic amines studied by Brüschweiler and Merlot (2017). This is a subset of the 397 aromatic amines as potential cleavage products from the total 180 azo dyes used for clothing textiles. The Ames test was conducted under the assumption that the TA98 and TA100 strains are often sufficient to detect mutagenicity of most aromatic amines (Bentzien et al. 2010, Harding et al. 2015). The Ames assay outcomes reported by the FSVO (Brüschweiler and Merlot 2017) and the predictions based on our 3D-SDAR PAA model are listed in Table 4. It is interesting to note that although 3D-SDAR was based on a fairly structurally limited training set, it was the only technique able to correctly classify the 4,4′-cyclohexane-1,1-diyldianiline as mutagenic and, respectively, achieve the highest sensitivity (see Table 5). In terms of specificity, all three approaches seemed to have generated a significant proportion of false positives. However, the somewhat significant level of concordance between these predictions (with Kendall τ_b coefficients ranging from 0.417 to 0.685) suggested that there might be a rational explanation for (at

least) some of the misclassified compounds. It was hypothesized that the presence of a 1,4-diaminobenzene moiety (found in 3 of the chemicals classified as false positives) or a 2-hydroxyethylamine group (present in 2 "false positives") might influence the mechanism by which these chemicals cause damage to DNA and that this altered mechanism may go undetected by the *Salmonella typhimurium* strains used

Table 5 Comparison of the performance of Lhasa Limited's Derek Nexus, Lhasa Limited's Sarah Nexus and 3D-SDAR to predict the two-strain Ames results

	Sarah Nexus	Derek Nexus	3D-SDAR
Accuracy	0.38	0.61	0.61
Sensitivity	0.75	0.75	1.00
Specificity	0.29	0.58	0.53
NPV	0.83	0.92	1.00
PPV	0.20	0.27	0.31
TP	3	3	4
TN	5	11	10
FP	12	8	9
FN	1	1	0

Table 4 Ames assay data for the experimental validation set of 23 PAA and predictions generated by 3D-SDAR. The non-mutagenic compounds are labeled with "−", while the mutagenic chemicals are labeled with "+"

CAS	Chemical name	2 Strain ames [31] exp	3D-SDAR
3282-99-3	4,4′-Cyclohexane-1,1-diyldianiline	+	+
615-47-4	1,2,4-Benzenetriamine	+	+
15791-87-4	4,6-Diamino-1,3-benzenediol	+	+
84-67-3	m-Tolidine	+	+
92-65-9	2-[(4-Aminophenyl)(ethyl)amino]ethanol	−	+
148-71-0[a]	1-(Diethylamino)-3-methylbenzene-4-amine	−	+
2359-51-5[b]	2-[(4-Amino-3-methylphenyl)(ethyl)amino]ethanol	−	+
105297-10-7	4-(1,1-Dioxido-4-thiomorpholinyl)aniline	−	−
23342-49-6	4-Amino-3-hydroxy-N-(2-methoxyphenyl)-2-naphthamide	−	+
52943-88-1	3-Methyl-1-phenyl-1H-pyrazole-4,5-diamine	−	+
88-50-6	4-Amino-2,5-dichlorobenzenesulfonic acid	−	−
98-44-2	1-Anilino-2,5-disulfonic acid	−	−
41772-23-0	2-Amino-1-naphthol	−	+
5856-00-8	N-(4-Amino-3-methylphenyl)-N-ethylbenzamide	−	+
2373-98-0	4,4′-Diamino-3,3′-biphenyldiol	−	+
117-62-4	2-Naphthylamine-1,5-disulfonic acid	−	−
137-51-9	4-Amino-1,3-benzenedisulfonic acid	−	−
6442-08-6	4,4′-Cyclohexane-1,1-diylbis(2-methylaniline)	−	−
98-36-2	2-Chloroaniline-5-sulfonic acid	−	−
2835-99-6	4-Amino-m-cresol	−	−
27746-11-8	(4-Amino-2-methylphenyl)dimethylamine	−	−
5272-86-6	3,5-Dimethyl-1H-pyrazol-4-amine	−	−
3204-61-3	1,2,4,5-Benzenetetramine	−	+

[a]Purchased as CAS 2051−79−8

[b]Purchased as CAS 25646−77−9

(Yoshida et al. 1998). This was the case with 2-[(4-aminophenyl)(ethyl)amino]ethanol, whose closest better-profiled analog 2,2′-[(4-aminophenyl)azanediyl]di(ethan-1-ol) demonstrated activity in the WP2 strain of *Escherichia coli*, but not in any of the *Salmonella typhimurium* strains.

Hence, it can be concluded that the discrepancy between the experimental and predicted mutagenicities could largely be attributed to the restricted set of strains used in the two-strain Ames test, as opposed to the broader sets, evidence from which is incorporated in Sarah Nexus, Derek and 3D-SDAR. For example, Sarah Nexus uses data from a 5 strain Ames test, whereas 3D-SDAR is based on a *P*-score, a component of which is the overall call.

Structural factors affecting the mutagenic potential of aromatic amines

The structural interpretation of both PAA and SAA models was based on the first two latent variables (i.e., those explaining most of the variance in mutagenicity data). On each of the 100 runs, the top 10 most positively and most negatively weighted bins for each of the two latent variables were extracted and their frequencies of occurrence (expressed as percentages) were calculated using the following formula: $F = 100 \times (\text{Count}/10 \times N_{LV} \times \text{Runs})$. In this formula, Count is the number of times a specific bin is found in the accumulated list of bins, N_{LV} is the number of latent variables and Runs represents the number of randomization cycles. In the case of PAA, 23 positively weighted bins with a frequency of occurrence of more than one percent and a cumulative frequency of occurrence of 81.03% were selected to identify the structural features associated with an increase in the mutagenic potential. Using the same cutoff of 1% resulted in the selection of 29 negatively weighted bins with a cumulative frequency of occurrence of 77.13%. In the case of SAA, the first 12 most frequently occurring bins with positive weights had a cumulative frequency of occurrence of 95.80%, whereas the cumulative frequency of occurrence of the 26 most negatively weighted bins was 81.40%. Due to the fewer number of both positively and negatively weighted individual bins needed to achieve a comparable or larger cumulative frequency of occurrence, in comparison to the PAA set, the SAA set is characterized by more clearly defined structural trends in the data. The most frequently occurring bins were further projected on the molecular structures to identify recurring patterns associated with an effective enhancement or reduction of the mutagenic potential of PAAs or SAAs. It is important to emphasize that the abundant atom specific information encoded in the 3D-SDAR fingerprints augmented with the robust bagging-like PLS algorithm allows for identification of "true" negative contributions that might be particularly difficult or often impossible to detect via alternative techniques. Such structural features are not simply lacking contribution to the overall mutagenic potential, but are known to effectively lower it when present in the structures of otherwise potential mutagens.

All structural features discussed below were found to be consistent with earlier (Q)SAR findings and expert knowledge derived fragments as well as with known mechanisms of mutagenicity (Ahlberg et al. 2016; Benfenati et al. 2015; Gadaleta et al. 2016).

Primary aromatic amines

The presence of several highly specific structural factors was found to enhance the mutagenic potential of PAAs. Aromatic amines containing conjugated planar π-systems such as fluorene or carbazole derivatives, including those with fused benzene rings such as naphthalene and anthracene were frequently present (select examples are shown in Fig. 5). These conjugated systems were often found to be long and narrow, with their width usually not exceeding two side-by-side benzene rings or 9 Å (taking into account the hydrogen atoms). Due to these highly specific geometric restrictions, the conjugated planar aromatic amines are likely exerting their effect through DNA intercalation leading to frame-shift mutations—a well-known mechanism enhancing mutagenicity (Ames et al. 1972). However, since the PAA data set consists only of aromatic amines, it is possible that the amino group is only a passenger, while the true "mutagenophore" is the planar conjugated aromatic system. More rigid polycyclic aromatic amines such as biphenylamine and especially their ortho substituted sterically hindered derivatives will exhibit a stronger mutagenic potential than those in which the aromatic rings are separated by two or more rotatable bonds (as in 2-benzylaniline). Another, strong "mutagenophore" contender was the nitrophenyl group (see Fig. 5), found among many PAAs and SAAs characterized by high *P*-scores. Its presence has been associated with various possible mechanisms of DNA damage including intercalation, adduct formation and in specific cases incorporation of the active molecule as a base analog into DNA during DNA synthesis (Matsuda et al. 1991).

The presence of the following structural factors was found to be associated with an overall reduction of the mutagenic potential: (1) ester, carboxyl and carboxamide groups directly attached to the phenylamine ring, with the ortho position being preferable (see Fig. 6 for select examples); (2) bulky (often aliphatic) substituents on both sides of the amino group; (3) diaminopyrimidine or diaminotriazine derivatives and (4) monoaromatic amines (except for the discussed above nitrophenylamines). In these cases, the reduction of mutagenic potential is likely caused by either the nearby bulky substituents blocking the free access to the nitrogen atom or its deactivation caused by redistribution of
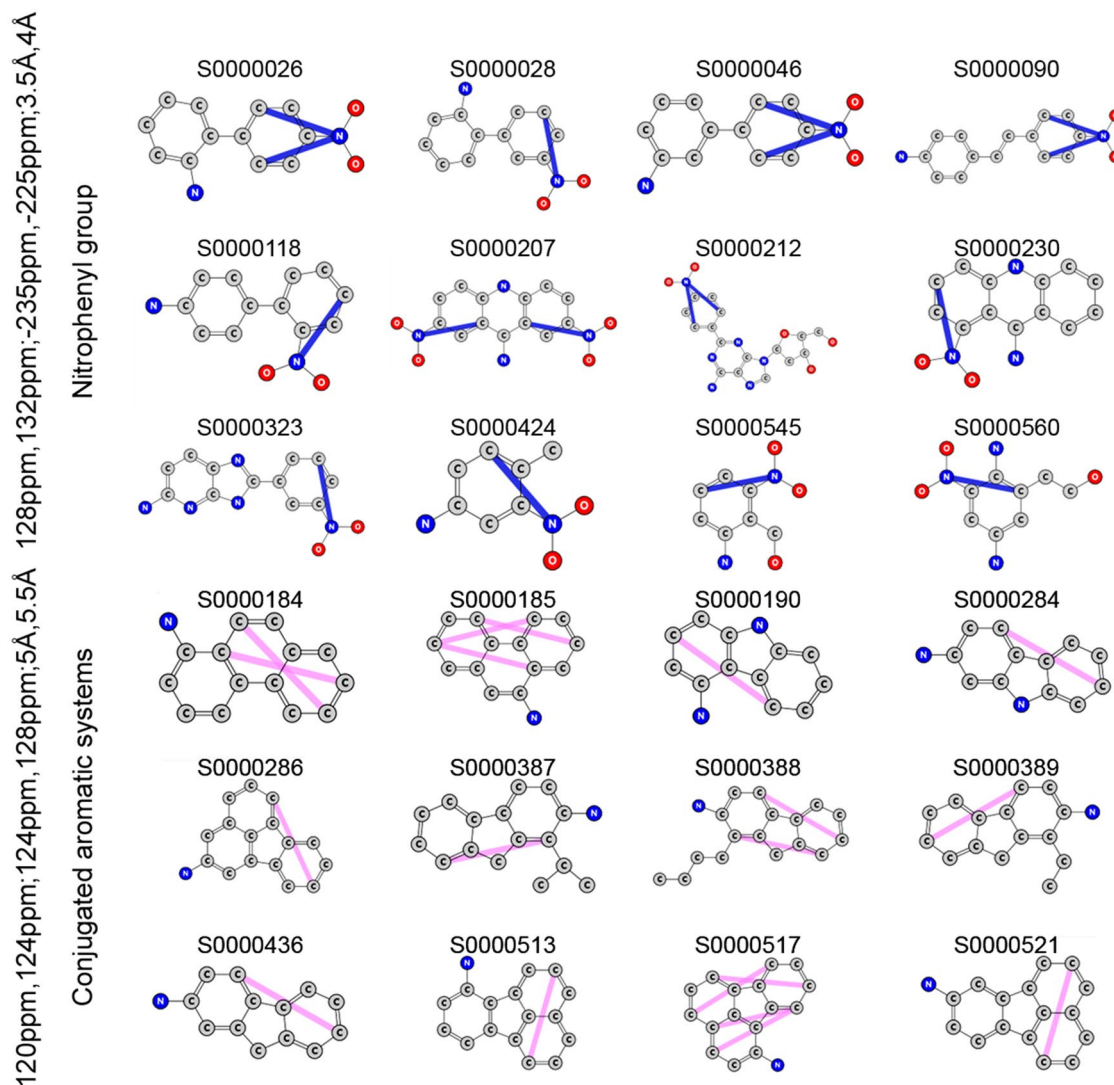
Fig. 5 Structural features associated with an increase in the mutagenic potential of PAAs

the electron density in the molecules and its concentration around the nearby more electronegative oxygens.

Secondary aromatic amines

In addition to the above-described structural features found to enhance the mutagenic potential of PAAs, specific patterns better expressed in the SAA data set confirmed indirectly one of the postulated mechanisms of aromatic amine activation in their path to forming DNA adducts. This is indeed the case of the phenylhydroxylamine derivatives (some examples of which are shown on Fig. 7) which undergo activation through conversion to their corresponding hydroxyl amines, further forming esters and finally producing nitrenium ions via elimination of the OR group. The nitrenium ions further form DNA adducts (Ford and Herman

1992). As with the PAAs, many SAAs containing either conjugated planar aromatic systems or somewhat more rigid (due to steric hindrance) polycyclic aromatic amine moieties were found to exhibit strong mutagenic potentials. One striking case was that all SAAs containing nitrothiophene moieties (see Fig. 7) were characterized by multiple overlaid positively weighted bins indicating a significant mutagenic effect. Although multiple positively weighted bins were also present in most of the cases discussed above (but only the most significant of them were visualized for simplicity), the large variety and recurring consistency of the bins associated with the presence of nitrothiophene moieties suggested that they carry an exceptionally high mutagenic potential. The DNA damaging potential of nitrothiophenes was noticed early on (Wang et al. 1975) with later studies demonstrating that their activation is largely dependent on the bacterial
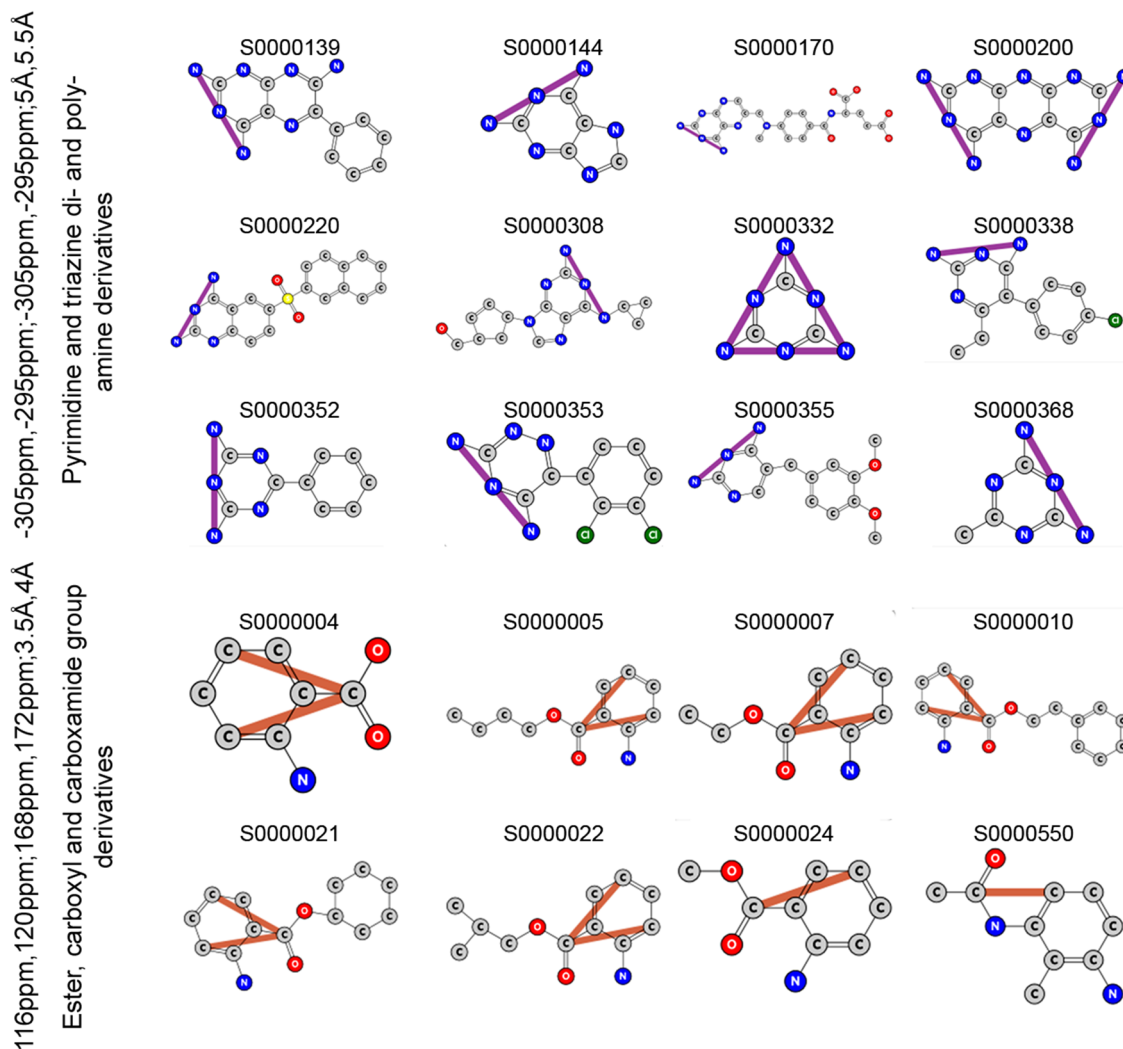
Fig. 6 Structural features associated with a decrease in the mutagenic potential of PAAs

nitroreductase, and therefore, they do not require a S9 fraction (Hrelia et al. 1990). Nitrothiophenes mutagenicity is likely caused by their reduction to diamagnetic and free radical intermediates further forming hydroxylamines (Hrelia et al. 1990).

Examination of the projection of the negatively weighted bins on the molecular structures of SAAs revealed several structural features that if present would effectively reduce the mutagenic potential of otherwise potentially active aromatic amines. Two such examples are the sulphonamide and acetamide residues (see Fig. 8), where the mechanism of deactivation is likely associated with the electron withdrawing effect on the nitrogen atom, thus hindering its oxidation in the conversion to a nitrenium ion. Similar to PAA, bulky substituents near the amino group, ester, carboxyl and carboxamide groups were also found to lower the mutagenic potential of SAA.

The structural interpretation of the 3D-SDAR models discussed in this section clearly demonstrates that the mutagenic potential of both PAAs and SAAs is driven by similar structural features. Although some are better expressed in one of the data sets and not in the other, this is likely a statistical artefact arising as a result of the differences in the coverage of the chemical space.

Conclusions

Using a carefully designed algorithm focused on validation and interpretability 3D-SDAR was able to successfully model diverse classes of primary and secondary aromatic amines. Experimental validation using 23 aromatic amines demonstrated predictive performance comparable to that of
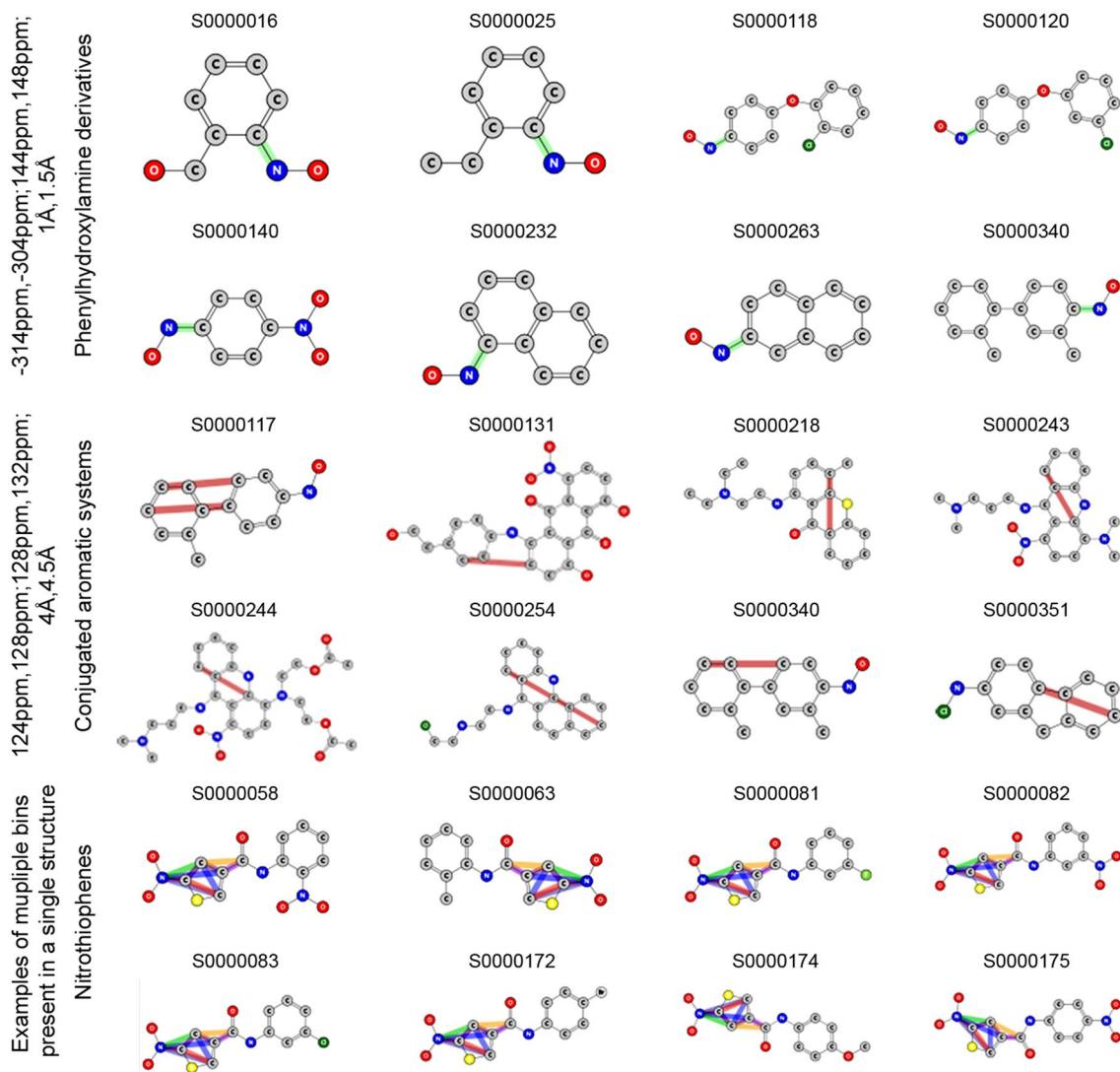
Fig. 7 Structural features associated with an increase in the mutagenic potential of SAAs

two widely used commercial systems developed by Lhasa Limited, namely Sarah Nexus and Derek Nexus.

The aggregated positively and negatively weighted 3D-SDAR bins and their projection on the standard coordinate space allowed the determination of structural features that either enhance or suppress the mutagenic potential of aromatic amines. Unlike most alternatives, 3D-SDAR was able to capture "true" negative contributions; i.e., functional groups or moieties whose presence actively reduces the overall mutagenic potential.

In compliance with the OECD requirements, our models were used to provide insights into the mechanisms by which the aromatic amines elicit their mutagenic effect. Observed structural trends in the data seem to confirm the postulated method of aromatic amines activation through conversion to their corresponding hydroxyl amines.
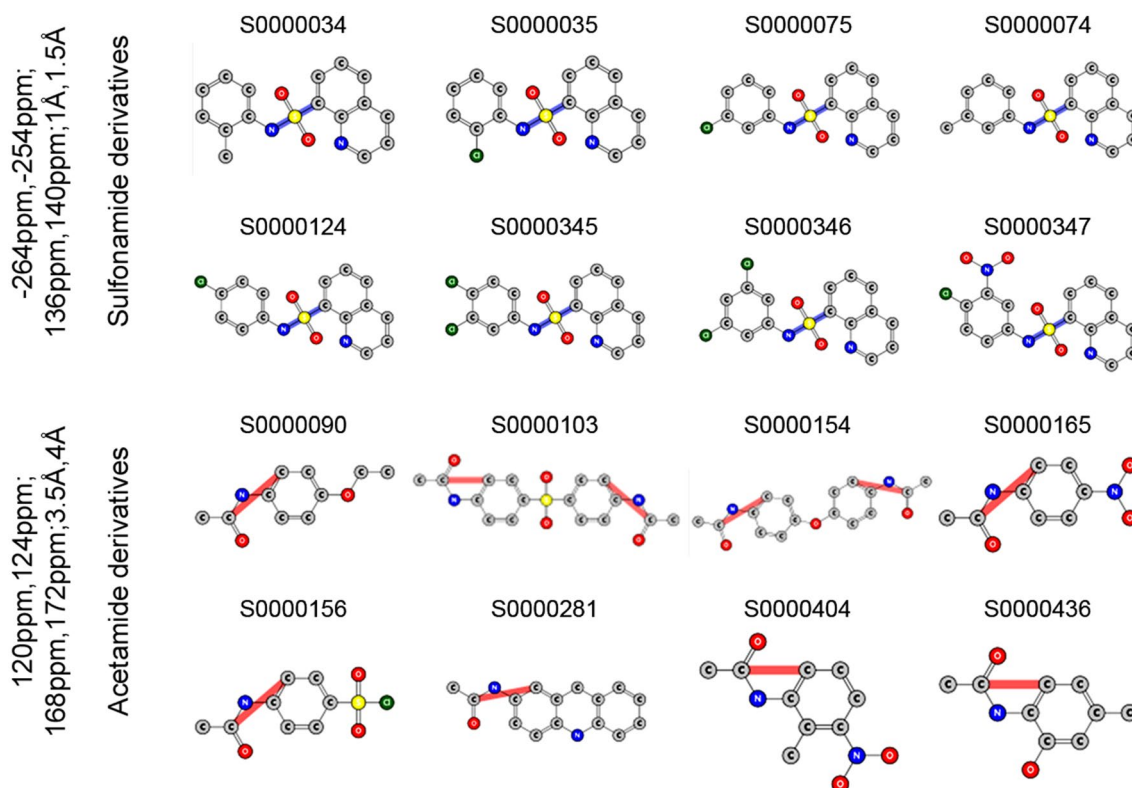
Fig. 8 Structural features associated with a decrease in the mutagenic potential of SAAs

Compliance with ethical standards

References

ACD/NMR (2011) Predictor Release 12.00, version 12.5. Advanced Chemistry Development, Toronto

ACD/XNMR Suite 12.00, version 12.5. (2011) Advanced Chemistry Development, Toronto

Ahlberg E, Amberg A, Beilke LD, Bower D, Cross KP, Custer L, Ford KA, Van Gompel J, Harvey J, Honma M, Jolly R, Joossens E, Kemper RA, Kenyon M, Kruhlak N, Kuhnke L, Leavitt P, Naven R, Neilan C, Quigley DP, Shuey D, Spirkl HP, Stavitskaya L, Teasdale A, White A, Wichard J, Zwickl C, Myatt GJ (2016) Extending (Q) SARs to incorporate proprietary knowledge for regulatory purposes: a case study using aromatic amine mutagenicity. Regul Toxicol Pharmacol 77:1–12. https://doi.org/10.1016/j.yrtph.2016.02.003

Ames BN, Gurney EG, Miller JA, Bartsch H (1972) Carcinogens as frameshift mutagens: metabolites and derivatives of 2-acetylaminofluorene and other aromatic amine carcinogens. Proc Natl Acad Sci USA 69(11):3128–3132

Ames BN, McCann J, Yamasaki E (1975) Methods for detecting carcinogens and mutagens with the Salmonella/mammalian-microsome mutagenicity test. Mutat Res/Environ Mutagen Relat Subj 31(6):347–364

Barber C, Cayley A, Hanser T, Harding A, Heghes C, Vessey JDJD., Werner S, Weiner SKSK., Wichard J, Giddings A, Glowienke S, Parenty A, Brigo A, Spirkl H-PH-PP., Amberg A, Kemper R, Greene N (2016) Evaluation of a statistics-based Ames mutagenicity QSAR model and interpretation of the results obtained. Regul Toxicol Pharmacol 76:7–20. https://doi.org/10.1016/j.yrtph.2015.12.006

Basak SC, Mills DR, Balaban AT, Gute BD (2001) Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: a hierarchical QSAR approach. J Chem Inf Comput Sci 41:671–678. https://doi.org/10.1021/ci000126f

Benfenati E, Manganelli S, Giordano S, Raitano G, Manganaro A (2015) Hierarchical rules for read-across and in silico models of mutagenicity. J Environ Sci Health C 33(4):385–403. https://doi.org/10.1080/10590501.2015.1096881

Benigni R, Bossa C (2011) Mechanisms of chemical carcinogenicity and mutagenicity: a review with implications for predictive toxicology. Chem Rev 111(4):2507–2536. https://doi.org/10.1021/cr100222q

Benigni R, Andreoli C, Giuliani A (1989) Interrelationships among carcinogenicity, mutagenicity, acute toxicity, and chemical structure in a genotoxicity data base. J Toxicol Environ Health Part A Current Issues 27.1:1–20. https://doi.org/10.1080/15287398909531275

Bentzien J, Hickey ER, Kemper RA, Brewer ML, Dyekjær JD, East SP, Whittaker M (2010) An in silico method for predicting Ames activities of primary aromatic amines by calculating the stabilities of nitrenium ions. J Chem Inf Model 50(2):274–297. https://doi.org/10.1021/ci900378x

Brüschweiler BJ, Merlot C (2017) Azo dyes in clothing textiles can be cleaved into a series of mutagenic aromatic amines which are not regulated yet. Regul Toxicol Pharmacol 88:214–226. https://doi.org/10.1016/j.yrtph.2017.06.012

Cariello NF, Wilson JD, Britt BH, Wedd DJ, Burlinson B, Gombar V (2002) Comparison of the computer programs DEREK and TOPKAT to predict bacterial mutagenicity. Mutagenesis 17(4):321–329. https://doi.org/10.1093/mutage/17.4.321

Casalegno M, Benfenati E, Sello G (2006) Application of a fragment-based model to the prediction of the genotoxicity of aromatic amines. Internet Electron J Mol Des 5(8):431–446. http://www.biochempress.com

Cash GG (2001) Prediction of the genotoxicity of aromatic and heteroaromatic amines using electrotopological state indices. Mutat Res 491:31–37. https://doi.org/10.1016/S1383-5718(00)00167-4

De Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. Chemom Intell Lab Systems 18:251–263. https://doi.org/10.1016/0169-7439(93)85002-X

Debnath AK, Debnath G, Shusterman AJ, Hansch C (1992) A QSAR investigation of the role of hydrophobicity in regulating mutagenicity in the Ames test: 1. Mutagenicity of aromatic and heteroaromatic amines in Salmonella typhimurium TA98 and TA100. Environ Mol Mutagen 19:37–52. https://doi.org/10.1002/em.2850190107

Fan M, Byrd C, Compadre CM, Compadre RL (1998) Comparison of CoMFA models for Salmonella Typhimurium TA98, TA100+ TA98+ S9 and TA100+ S9 mutagenicity of nitroaromatics. SAR QSAR Environ Res 9(3–4):187–215. https://doi.org/10.1080/10629369808039156

Ford GP, Herman PS (1992) Relative stabilities of nitrenium ions derived from polycyclic aromatic amines. Relationship to mutagenicity. Chem Biol Interact 81(1–2):1–18. https://doi.org/10.1016/0009-2797(92)90023-E

Gadaleta D, Manganelli S, Manganaro A, Porta N, Benfenati E (2016) A knowledge-based expert rule system for predicting mutagenicity (Ames test) of aromatic amines and azo compounds. Toxicol 370:20–30. https://doi.org/10.1016/j.tox.2016.09.008

Gramatica P, Consonni V, Pavan M (2003) Prediction of aromatic amines mutagenicity from theoretical molecular descriptors. SAR QSAR Environ Res 14:237–250. https://doi.org/10.1080/1062936032000101484

Harding AP, Popelier PL, Harvey J, Giddings A, Foster G, Kranz M (2015) Evaluation of aromatic amines with different purities and different solvent vehicles in the Ames test. Regul Toxicol Pharmacol 71(2):244–250. https://doi.org/10.1016/j.yrtph.2014.12.009

Hatch FT, Colvin ME (1997) Quantitative structure-activity (QSAR) relationships of mutagenic aromatic and heterocyclic amines. Mutat Res 376:87–96. https://doi.org/10.1016/S0027-5107(97)00029-8

Helma C (2006) Lazy structure-activity relationships (lazar) for the prediction of rodent carcinogenicity and Salmonella mutagenicity. Mol Divers 10:147–158. https://doi.org/10.1007/s11030-005-9001-5

Hrelia P, Morotti M, Scotti M, Paolini M, Spinelli D, Cantelli FG (1990) The Direct-Acting mutagenicity of nitrothiophenes derivatives in Salmonella Typhimurium. In: Howard PC, Hecht SS, Beland FA (eds) Nitroarenes. Environ Sci Res, vol 40. Springer, Boston, pp 320–320. https://doi.org/10.1007/978-1-4615-3800-4_32

HyperChem 8 Professional, version 8.03. 2007. HyperCube Inc., Gainesville, FL

ICH M7 (2015) http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM347725

Leong MK, Lin SW, Chen HB, Tsai FY (2010) Predicting mutagenicity of aromatic amines by various machine learning approaches. Toxicol Sci 116:498–513. https://doi.org/10.1093/toxsci/kfq159

Manganelli S, Benfenati E, Manganaro A, Kulkarni S, Barton-Maclaren TS, Honma M (2016) New quantitative structure–activity relationship models improve predictability of Ames mutagenicity for aromatic azo compounds. Toxicol Sci 153(2):316–326. https://doi.org/10.1093/toxsci/kfw125

Maran U, Karelson M, Katritzky AR (1999) A comprehensive QSAR treatment of the genotoxicity of heteroaromatic and aromatic amines. Quant Struct-Act Relat 18:3–10. (10.1002/(SICI)1521-3838(199901)18:1<03::AID-QSAR3>3.0.CO;2-P)

Marchant CA, Briggs KA, Long A (2008) In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic. Toxicol Mech Methods 18(2–3):177–187. https://doi.org/10.1080/15376510701857320

Martin T (2016) User's guide for T.E.S.T. (version 4.2) (Toxicity Estimation Software Tool) A program to estimate toxicity from molecular structure. US EPA Office of Research and Development, Washington, DC, EPA/600/R-16/058

MATLAB (2012) Version 8.0. The MathWorks, Cambridge

Matsuda A, Akashi M, Ohara Y, Wataya Y, Hayatsu H, Ueda T (1991) Mutagenicity of (p-nitrophenyl) adenines in Salmonella typhimurium. Mutat Res Lett 263(2):93–100. https://doi.org/10.1016/0165-7992(91)90065-C

McCarren P, Bebernitz GR, Gedeck P, Glowienke S, Grondine MS, Kirman LC, Klickstein J, Schuster HF, Whitehead L (2011) Avoidance of the Ames test liability for aryl-amines via computation. Bioorg Med Chem 19(10):3173–3182. https://doi.org/10.1016/j.bmc.2011.03.066

Mccarren P, Springer C, Whitehead L (2011) An investigation into pharmaceutically relevant mutagenicity data and the influence on Ames predictive potential J. Cheminform 3:51. https://doi.org/10.1186/1758-2946-3-51

OECD, The report from the Expert Group on (Quantitative) Structure-ActiVity Relationship (Q)SAR on the Principles for the Validation of (Q)SARs (2004) OECD series on Testing and Assessment No. 49, Organisation for Economic Cooperation and Development, Paris, France

Patlewicz G, Jeliazkova N, Safford RJ, Worth AP, Aleksiev B (2008a) An evaluation of the implementation of the Cramer classification scheme in the Toxtree software. SAR QSAR Environ Res 19(5–6):495–524. https://doi.org/10.1080/10629360802083871

Patlewicz G, Jeliazkova N, Gallegos Saliner A, Worth AP (2008b) Toxmatch—a new software tool to aid in the development and evaluation of chemically similar groups. SAR QSAR Environ Res 19:397–412. https://doi.org/10.1080/10629360802083848

Saiakhov RD, Klopman G (2008) MultiCASE expert systems and the reach initiative. toxicol. Mech Methods 18:159–175. https://doi.org/10.1080/15376510701857460

Slavov S, Beger RD (2016) Rigorous 3D-SDAR modeling strategy for ToxCast estrogen receptor data classification, validation and feature extraction. Env Toxicol Chem 36(3):823–830. https://doi.org/10.1002/etc.3578

Slavov S, Wilkes J, Buzatu D, Kruhlak N, Willard J, Hanig J, Beger R (2014) Computational identification of a phospholipidosis toxicophore using ^{13}C and ^{15}N NMR-distance based fingerprints. Bioorg Med Chem 22:6706–6714. https://doi.org/10.1016/j.bmc.2014.08.021

Slavov S, Stoyanova-Slavova I, Li S, Zhao J, Huang R, Xia M, Beger R (2017) Why are most phospholipidosis inducers also hERG. Blockers?. Arch Toxicol 91(12):3885–3895. https://doi.org/10.1007/s00204-017-1995-9

Stewart BW, Wild CP, World Cancer Report (2014), International Agency for Research on Cancer, WHO (2014) Retrieved June 13, 2016, from http://www.thehealthwell.info/node/725845

Toropov AA, Toropova AP (2001) Prediction of heteroaromatic amine mutagenicity by means of correlation weighting of atomic orbital graphs of local invariants. J Mol Struct Theochem 538:287–293. https://doi.org/10.1016/S0166-1280(00)00713-2

Torres-Cartas S, Martın-Biosca Y, Villanueva-Camanas RM, Sagrado S, Medina-Hernandez MJ (2007) Biopartitioningmicellar chromatography to predict mutagenicity of aromatic amines. Eur J Med Chem 42:1396–1402. https://doi.org/10.1016/j.ejmech.2007.02.022

Vracko M, Mills D, Basak SC (2004) Structure-mutagenicity modelling using counter propagation neural networks. Environ Toxicol Pharmacol 16:25–36. https://doi.org/10.1016/j.etap.2003.09.004

Wang CY, Muraoka K, Bryan GT (1975) Mutagenicity of nitrofurans, nitrothiophenes, nitropyrroles, nitroimidazole, aminothiophenes, and aminothiazoles in Salmonella typhimurium. Cancer Res 35(12):3611–3617

Winder C, Stacey NH (eds). (2004) Occupational toxicology. CRC press, Boca Raton, p 242

Yang C, Hasselgren CH, Boyer S, Arvidson K, Aveston S, Dierkes P, Benigni R, Benz RD, Contrera J, Kruhlak NL, Matthews EJ, Han X, Jaworska J, Kemper RA, Rathman JF, Richard AM (2008) Understanding genetic toxicity through data mining: the process of building knowledge by integrating multiple genetic toxicity databases. Toxicol Mech Methods 18:277–295. https://doi.org/10.1080/15376510701857502

Yoshida R, Oikawa S, Ogawa Y, Miyakoshi Y, Ooida M, Asanuma K, Shimizu H (1998) Mutagenicity of p-aminophenol in E. coli WP2uvrA/pKM101 and its relevance to oxidative DNA damage. Mutat Res 415(1):139–150. http://www.biomedsearch.com/nih/Mutagenicity-p-aminophenol-in-E/9711270.html