Published in final edited form as:

SAR QSAR Environ Res. 2012 October ; 23(7-8): 775–795. doi:10.1080/1062936X.2012.728996.

Chemical structure determines target organ carcinogenesis in rats

C. A. Carrasquer¹, N. Malik¹, G. States¹, S. Qamar¹, S.L. Cunningham¹, and A.R. Cunningham^{1,2,3,*}

¹James Graham Brown Cancer Center, University of Louisville, Louisville, KY 40202

²Department of Medicine, University of Louisville, Louisville, KY 40202

³Department of Pharmacology and Toxicology, University of Louisville, Louisville, KY 40202

Abstract

SAR models were developed for 12 rat tumour sites using data derived from the Carcinogenic Potency Database. Essentially, the models fall into two categories: Target Site Carcinogen – Non-Carcinogen (TSC-NC) and Target Site Carcinogen – Non-Target Site Carcinogen (TSC-NTSC). The TSC-NC models were composed of active chemicals that were carcinogenic to a specific target site and inactive ones that were whole animal non-carcinogens. On the other hand, the TSC-NTSC models used an inactive category also composed of carcinogens but to any/all other sites but the target site. Leave one out validations produced an overall average concordance value for all 12 models of 0.77 for the TSC-NC models and 0.73 for the TSC-NTSC models.

Overall, these findings suggest that while the TSC-NC models are able to distinguish between carcinogens and non-carcinogens, the TSC-NTSC models are identifying structural attributes that associate carcinogens to specific tumour sites. Since the TSC-NTSC models are composed of active and inactive compounds that are genotoxic and non-genotoxic carcinogens, the TSC-NTSC models may be capable of deciphering non-genotoxic mechanisms of carcinogenesis. Together, models of this type may also prove useful in anticancer drug development since they essentially contain chemicals moieties that target specific tumour site.

Keywords

structure-activity relationship; cancer; organ-specific carcinogens; genotoxic carcinogens; non-genotoxic carcinogens

1. Introduction

A number of methods exist to determine the carcinogenicity of chemicals ranging from short-term *in vitro* genotoxicity tests [1] to whole animal carcinogenesis bioassays [2] to epidemiological studies [3]. Using data generated from these short-term tests and animal bioassays, a number of methods have also been developed, that to varying degrees of success, are capable of predicting the carcinogenic potential of chemicals. For example, we have reported structure-activity relationship (SAR) models for chemical carcinogenesis in mice [4] and rats [5] derived using the CASE/MULTICASE (MCASE) expert system with data derived from the Carcinogenic Potency Database (CPDB) [6]. From these studies, the

*Corresponding author. al.cunningham@louisville.edu.

Supporting Information Available

best rat and mouse models had a concordance between experimental and SAR-predicted values of 71 and 78%, sensitivity of 69 and 77%, and specificity of 73 and 78%, respectively [4, 5]. MCASE MC4PC and MDL-QSAR models have also been developed by Contrera et al., using 1540 compounds tested for rodent carcinogenicity that were compiled by the Food and Drug Administration (FDA). Concordance values of 66 and 69%, sensitivity of 61 and 63%, and specificity of 71 and 75%, were reported respectively for MCASE MC4PC and MDL-QSAR [7]. The utility and application of other toxicologically-focused predictive methods has been reviewed in-depth [8–12].

Since surrogate tests and carcinogen animal bioassays themselves are not reproducible with 100% concordance, the SAR models derived from these data cannot be expected to be completely accurate. For instance, the NTP's Salmonella mutagenicity database, which is derived from a standardized protocol, has been estimated to be about 85% reproducible [13]. Moreover, it was found that based on “near-replicate” experiments in the CPDB, there was also a degree of non-reproducibility [14, 15]. For example, 43 out of 54 chemicals tested in similar experiments for their ability to induce cancer in mice were concordant (i.e., 80% reproducible) and 88 out of 104 chemicals tested for cancer in rats were concordant (i.e., 85% reproducible) [15]. Moreover, Gottman et al., using the CPDB found only a 57% concordance between 121 compounds tested by the NTP/NCI where literature values were also available [16].

Furthermore, *in silico* models cannot be more predictive than the original *in vitro* or *in vivo* data is reproducible. To address the general accuracy of some popular predictive techniques for carcinogenicity, Benigni and Bossa reported that while internal validation methods along the line of LOO may overestimate the accuracy of the models, true external validation correct prediction rates may be between 70 and 100% [17]. In fact, in a recent investigation by Milan et al. the predictivity of four QSAR expert systems for carcinogenesis (CAESAR, TOPKAT, LAZAR and MultiCASE) was determined using a pure external test set [18]. Although the internal validation of the training sets produced sensitivities between 57 and 89%, specificities between 69 and 89% and accuracies between 64 and 89%, prediction of the external test compounds produced sensitivities between 43 and 72%, specificities between 41 and 78% and accuracies between 57 and 67%. They speculate that the poor performance was due to the uncertainty and variability of the animal studies on carcinogenicity, which are quite old and completed with inconsistent experimental procedures [18].

The CASE/MULTICASE SAR models of rat and mouse carcinogens developed by us, while being predictive, also provided insight into the structural underpinnings for species-specific carcinogenesis. Many, though not all, of the readily explainable attributes of these models corresponded with the genotoxic or electrophilic paradigm of carcinogenesis [19]. This is not surprising given the large numbers of electrophilic or proelectrophilic carcinogens used to build the models. In fact, in a recent review article by Benigni and colleagues, it was noted that there currently is no mutagenicity based tools for assessing the cancer risk for non-DNA reactive agents [20]

When considering the specific role that genotoxicants play in tumor site specific carcinogenesis, results from the comet (i.e., alkaline single cell gel electrophoresis) assay are interesting since the method can detect organ-specific *in vivo* chemical-induced genotoxicity. When applying this assay to a set of 208 chemicals previously tested for carcinogenicity, Sasaki *et al.* found that many of the organs that displayed DNA damage were not necessarily targets for carcinogenicity but nearly all organs displaying carcinogenicity were also targets of genotoxicity [21]. They concluded that although genotoxicity is generally necessary for carcinogenicity, it is not a sufficient predictor of

organ-specific carcinogenicity [21, 22]. In other words, although genotoxicity is a mechanistic link to cancer, DNA adducts can be found in similar levels between cancer target and non-target organs [23]. In fact, when considering the analysis of *in vivo* biomarkers, DNA from an often non-cancer target organ is used as a surrogate for other sites that might be the target for carcinogenesis [24]. When discussing organ-specific carcinogenicity,

For a specific example, acrylamide has gained attention partly due to its presence in cooked [25, 26] and baby [27] food. Acrylamide is a rat carcinogen [6] and although it is not a *Salmonella* mutagen, it is genotoxic in other assays (reviewed in [28]). By dosing mice and rats with acrylamide Doerge et al. observed adduct formation in carcinogenic target and non-target organs [29]. Manière reports that the organ-specific carcinogenesis of acrylamide was not explained by the formation of adducts in the target organ since adducts were uniformly produced in target and non-target organs [30]. And recently, Lafferty observed an increase in DNA synthesis that did correlate to acrylamide's carcinogenic targets and also observed that oxidative metabolism or the oxidized metabolites (i.e., DNA damaging events) did not seem to account for the observed increase in DNA synthesis [28].

Since whole-animal carcinogenicity data deals with many underlying and competing mechanisms, the development of organ-specific carcinogenicity SAR models is appealing. The FDA's National Center for Toxicological Research noted that FDA reviewers are interested in organ-specific carcinogenicity to aid in evaluating new chemicals [31]. In their preliminary SAR analyses of liver carcinogens, they obtained an overall predictability of 63%, with a sensitivity of 30% and a specificity of 77% [31].

Recently, based on rat mammary carcinogens reported in the CPDB [32] we used the cat-SAR expert system to analyse mammary-specific carcinogenesis [33]. Cat-SAR models are built through a comparison of structural features found amongst categorized compounds in the model's learning set. Generically, these categories are biologically active and inactive compounds. When just considering whole animal carcinogenesis, the categories are carcinogens and non-carcinogens. However, when considering organ-specific carcinogenesis, the question arises as to the selection of the inactive or non-carcinogenic compounds and two options can be considered. One option pits carcinogens to one organ against non-carcinogens and the other pits them to carcinogens at all other sites. In the case of mammary carcinogens two unique SAR models were developed: A mammary carcinogen –non carcinogen (MC-NC) model and a mammary carcinogen - non-mammary carcinogen (MC-NMC) model. Thus the MC-NMC model considered carcinogens as both active (i.e., mammary carcinogens) and inactive (i.e., carcinogens to other sites). Analysis of this model therefore did not classify carcinogens from non-carcinogens, but rather it classified mammary carcinogens from other carcinogenic compounds based on structural difference between the two sets of carcinogens. For clarity, organ-specific carcinogens are not necessarily specific to one target site but rather the term organ-specific indicates that a carcinogen has been determined to specifically induce cancer at a certain organ or organs. For example, acrylamide specifically induces cancer in the rat mammary gland in addition to three other specific sites included in this study (clitoral gland, nervous system, and uterus).

Based on a leave-one-out validation, the best rat MC-NC model achieved a concordance between experimental and predicted values of 84%, a sensitivity of 79%, and a specificity of 89%. The best rat MC-MNC model achieved a concordance of 78%, with a sensitivity of 82%, and a specificity of 74%. Therefore, the MC-NMC model identified structural attributes that are useful in addressing the question of “why do some carcinogens cause breast cancer” which is a different question than “why do some chemicals cause cancer”. In

essence, a tiered computational approach could be used to first identify carcinogens, and then identify those that might be active at the mammary gland.

Described herein, we extended our analyses to a set of 12 rat tumour target sites. As with the MC-NMC model, SAR analyses were conducted on learning sets composed of two activity categories (i.e., carcinogenic and non-carcinogenic). The carcinogen category contained chemicals that were carcinogenic to a specified target site and the non-carcinogen category contained 1) chemicals that were non-carcinogens and 2) chemicals that were carcinogens at sites other than the target site. For simplicity, the active category will be referred to as Target Site Carcinogen – Non-Carcinogens (TSC-NC) and Target Site Carcinogen -Non-Target Site Carcinogens (TSC-NTSC). Moreover, since the TSC-NTSC models are composed of carcinogens as both active and inactive compounds, and by extension the active and inactive categories are composed of genotoxicants, the TSC-NTSC models may be capable of deciphering the non-genotoxic mechanisms of carcinogenesis.

2. Materials and methods

2.1 Learning sets

The CPDB standardizes the experimental results (whether positive or negative for carcinogenicity), including qualitative data on strain, sex, route of compound administration, target organ, histopathology, and the author's opinion and reference to the published paper, as well as quantitative data on carcinogenic potency, statistical significance, tumour incidence, dose-response curve shape, length of experiment, duration of dosing, and dose rate [34]. Moreover, a potency value for carcinogens, the TD₅₀ is also available. The TD₅₀ is defined as "that dose rate (in mg/kg body weight/day) which, if administered chronically for the standard lifespan of the species, will halve the probability of remaining tumourless throughout that period" [34].

For this study, we used the rat data from the CPDB summary by target organ list published on the CPDB's website [6] to develop SAR models for 12 different tumour sites (i.e., clitoral gland, oesophagus, hematopoietic system, kidney, large and small intestines, liver lung, mammary gland, nasal cavity, nervous system, and uterus).

For each tumour site two sets of models were produced; a TSC-NC model composed of carcinogens to the target site and non-carcinogens and a TSC-NTSC model composed of carcinogens to the target site and carcinogens to all other sites. Furthermore since there were ample non-carcinogens and non-target site carcinogens for the selected sites, each TSC set was matched with three random and equal in number sets of inactive compounds.

The cat-SAR learning set consisted of the chemical name, its structure as a .MOL2 file, and its categorical designation (e.g., one or zero). Organic salts were included as the freebase. Simple mixtures and technical grade preparations were included as the major or active component. Metals, metalloorganic compounds, polymers, and mixtures were not included.

2.2 In silico chemical fragmentation and the compound-fragment data matrix

Using the Tripos Sybyl HQSAR module, each chemical was fragmented *in silico* into all possible fragments meeting user-specified criteria. HQSAR allows the user to select attributes for fragment determination including atom count (i.e., size of the fragment), bond types, atomic connections (i.e., the arrangement of atoms in the fragment), hydrogen atoms, chirality, and hydrogen bond donor and acceptor groups. Fragments can be linear, branched, or cyclic moieties. Models developed herein contained fragments between two and seven atoms in size and considered atoms, bond types, and atomic connections, and explicate hydrogen atoms.

Upon completion of the fragmentation routine, a Sybyl HQSAR add-on procedure produces the compound-fragment data matrix as a text file. In the matrix, the rows are intact chemicals and the columns are molecular fragments. Thus for each chemical, a tabulation of all its fragments is recorded across the table rows and for each fragment all chemicals that contain it are tabulated in each column.

The HQSAR module is not used for statistical analysis or model development. The compound-fragment matrix is then analysed, using the cat-SAR program to identify structural features associated with active and inactive compounds, validate the model, and predict the activity of untested compounds. The cat-SAR program, learning sets and the compound-fragments matrix are available through the corresponding author and the mammary carcinogen models are also available in the Supporting Information.

2.3 Identifying “important” fragments of activity and inactivity

A measure of each fragment’s association with biological activity was next determined. To ascertain an association between each fragment and activity (or inactivity), a set of rules is established to choose “important” active and inactive fragments. The first selection rule is the number of times a fragment is identified in the learning set. For this exercise, it was set at between two and five compounds. For this parameter, we reasoned that by looking at fragments that came from between two and five compounds in the learning set, models derived in the two to three range would be more inclusive (i.e., higher coverage) while those in the four to five range would be more accurate (i.e., higher concordance). We note again, that in previous cat-SAR analyses the fragment number was arbitrarily set to three.

The second rule relates to the proportion of active or inactive compounds that contribute to each fragment. The proportion of active (or inactive) compounds associated with a particular fragment derived ranged from between 50% to 95%. We reasoned that even if a particular fragment is associated with activity, there may be other reasons (i.e., fragments) for its being inactive, thus it would not be expected to be found in 100% of the active compounds. A similar argument can be made for inactive fragments. Thus, by considering fragments toward the lower high end of the proportion scale (e.g., derived from 60% active and 40% inactive) we would expect models to again be more inclusive (i.e., higher coverage) while those derived from the higher end of the proportion scale (e.g., 90% active and 10% inactive) would be more accurate (i.e., higher concordance).

The compound learning set and list of “important fragments” used for model validation will be provided upon request by the corresponding author.

2.4 Rule optimization

Previous cat-SAR models required a relatively arbitrary parameter set that selected important fragments (fragment compound counts and fragment activity proportion values). For these analyses a rule optimization routine was employed. The optimization routine in this instance allowed the Number Rule to range between 1 and 8 and the Proportion Rule to range between 0.50 and 0.95. LOO validations were then conducted for each model and final models were selected that were both highly accurate (i.e., had a high concordance between experimental and predicted values) and highly predictive (i.e., made predictions on >90% of the chemicals in the learning set).

2.5 Model validation

A leave-none-out (LNO) and two cross-validation routines (i.e., leave-one-out (LOO) and leave-many-out (LMO)) were conducted for each model. For the LOO cross-validation, each chemical, one at a time, was removed from the total fragment set, and the n-1 model was

derived. Using the same criteria described above, the activity of the removed chemical was then predicted using the n-1 model. Predicted vs. experimental values for each chemical were then compared and the model's concordance, sensitivity, and specificity were determined.

For the LMO cross-validation, randomly selected sets of 10% of the chemicals (i.e., 20 chemicals in a learning set of 200) were removed from the total fragment set, and the n-10% model was derived. Again, the activity of each of the removed chemicals was then predicted using the n-10% model. This process was repeated 1000 times and the average predicted vs. experimental values for the chemicals in the left out sets were calculated and the model's concordance, sensitivity, and specificity were determined. It should be noted that the values in Table 1 and 2 for the LMO validations are non-integers since the process averaged a large number of individual runs wherein the number of active, inactive, and non-predicted compounds varied from set to set.

Cat-SAR predictions are based on two separate fragment sets (i.e., the active fragments and the inactive ones) and the predicted activity of a chemical is based on the average probability of all the active and inactive compounds contributing to its fragments. To best classify compounds back to an active or inactive category, we have adapted a routine from our previous MultiCASE work in which we identify an optimal cut-off point that best separates the probabilistic prediction of active and inactive compounds based on the drop-one validations.

2.6 Predicting activity

Once a final model is selected, the resulting list of fragments can then be used for mechanistic analysis, or to predict the activity of an unknown compound. In the latter circumstance, the cat-SAR program determines which, if any, fragments from the model's pool of significant fragments the test compound contains. If none are present, no prediction of activity is made for the compound. If one or more fragments are present, the number of active and inactive compounds containing each fragment is determined. The probability of activity or inactivity is then calculated based on the total number of active and inactive compounds that went into deriving each of the fragments.

The probability of activity was calculated with the cat-SAR FragSum routine. This method calculates the average probability of the active and inactive fragments contained in each compound it and is weighted to the number of active and inactive compounds that go into deriving each fragment. For example, if a compound contains two fragments, one being found in 9/10 active compounds in the learning set (i.e., 90% active) and the other being found in 3/3 inactive compounds (i.e., 0% active), the unknown compound will be predicted to be have a probability of activity of 69% (i.e., $9/10 \text{ actives} + 0/3 \text{ actives} = 9/13 \text{ actives}$ or 69% chance of activity).

3. Results and discussion

For each tumour site specific endpoint, three random sets of non-carcinogens (TSC-NC models) or non-target site carcinogens (TSC-NTSC models) were selected and rule-optimized final models were developed. LOO validations were used to select final models based on the following conditions: At least 90% of the compounds had to be predicted during the LOO validation and resulting sensitivity and specificity values had to be ≥ 0.60 . Two models were then selected: the model with the highest overall concordance and the model with near equal sensitivity and specificity (balanced) (see Table 2).

Of the 12 TSC-NC models, the oesophagus model had the highest concordance between experimental and predicted results. The average concordance for the three balanced models was 0.93 with the range of the three models between 0.91 and 0.94. The oesophagus TSC-NC best concordance models had an average concordance of 0.95 and ranged between 0.94 and 0.96. For these models the sensitivity and specificity values ranged between 0.89 and 1.00. The LMO cross-validation resulted in a concordance value of 0.91 and the LNO returned a concordance value of 0.94. On the other hand, the least predictive model was the kidney TSC-NC. The average concordance for the best concordance and balanced models was 0.66 and 0.67, respectively. The LMO cross-validation returned a concordance value of 0.66 and the LNO value was 1.00 (Table 1).

Considering the concordance values of the 12 models individually, one model (i.e., as mentioned the oesophagus) had a concordance value greater than 0.90, eight had values in the 0.80s (i.e., clitoral gland, large intestine, lung, mammary gland, nasal cavity, nervous system, small intestine and uterus), two in the 0.70s (i.e., hematopoietic system and liver), and one in the 0.60s (i.e., kidney). Standard deviation of concordance values for the best models ranged between 0.01 and 0.06 and for the balanced models they were between 0.01 and 0.09.

Of the 12 TSC-NTSC models, the oesophagus TSC-NTSC balanced model had the best average concordance of 0.86 with a range between 0.83 and 0.88. The oesophagus best concordance model had an average concordance of 0.87 and ranged between 0.85 and 0.88. The LMO cross-validation returned a concordance value of 0.86 and the LNO was 0.93. And again, the model with the lowest average concordance value was the kidney TSC-NTSC with average concordance values of 0.66 and 0.67 for the balanced and best concordance models, respectively. The LMO concordance value for the kidney TSC-NTSC was 0.64 and the LNO was 0.91.

Considering the individual concordance values for these models, three models had a concordance value in the 0.80s (i.e., clitoral gland, oesophagus, and small intestine), seven in the 0.70s (i.e., liver, lung, large intestine, mammary gland, nasal cavity, nervous system, and uterus), and two in the 0.60s (i.e., hematopoietic system and kidney). The standard deviation of concordance values ranged between 0.01 and 0.12 and 0.01 and 0.14 for the balanced and best concordance models, respectively.

Overall 36 individual TSC-NC and TSC-NTSC models (i.e., three random sets of inactive compounds for each target site) were developed. For each group of models only little to moderate deviations were observed, thus indicating that the accurate predictions made by the models were not spurious events based on a fortuitous selection of “good” compounds. This provides a degree of assurance that the models are in fact capable of distinguishing target site specific carcinogens based on the structural differences between them and other carcinogens (i.e., the TSC-NTSC models) and non-carcinogens (i.e., the TSC-NC models).

In a separate cat-SAR analysis of a set of rat carcinogens derived from the CPDB consisting of 531 carcinogens (to all sites) and 393 non-carcinogens, balanced and best model LOO validations returned concordance, specificity, and sensitivity values of 0.70 (unpublished results). Comparing the average concordance values of the 12 TSC-NC best models to the rat cancer model, 11 of the 12 were superior, with the kidney being the exception. When considering the three individual models for each endpoint (i.e., 36 models) only the three kidney models had concordance values less than 0.70. Moreover, the average concordance of the 36 separate best concordance models was 0.84 noting that Gold et al., estimated the reproducibility of the rat bioassay data at 0.85% [15].

The TSC-NTSC models in all cases have close but somewhat lower concordance values than for the TSC-NC models. This is not unexpected since a similar and modest difference was observed in our previous analysis of rat mammary carcinogens [33]. The high concordance values observed in the TSC-NTSC models warrant special notice since these models are constructed from only carcinogens (as active and inactive compounds). Hence these models are capable of classifying carcinogens based on their target site of activity.

These observations, based on the analysis of 12 separate organ specific models, suggest that the development and use of small organ-defined SAR can be more predictive and insightful than models built for species specificity.

For example, of the 12 target sites considered, PhIP is carcinogenic to four (i.e., the hematopoietic system, large and small intestines, and the mammary gland). Of the four TSC-NC models, PhIP was accurately predicted to be a carcinogen at that site (Table 3). Considering the TSC-NTSC models, PhIP was accurately predicted in the large and small intestines and mammary gland model and only incorrectly predicted as a non-hematopoietic system carcinogen in the hematopoietic TSC-NTSC model (Table 3). By chance, PhIP was also selected to be a non-target site carcinogen in the liver TSC-NTSC model where it was inaccurately predicted to be a liver carcinogen. Overall, based on a weight-of-evidence approach, PhIP would be estimated to be a carcinogen, and most of its target sites would be accurately determined.

In a less complicated example, C.I direct black 38 is solely a liver carcinogen and was correctly predicted to be so by both the TSC-NC and TSC-NTSC liver models (Table 3). By chance, the compound was also selected to be a non-target site carcinogen in the oesophagus, kidney, lung, and mammary gland models and it was correctly predicted to not be a carcinogen at all of those sites (Table 3). Similarly, 4-aminobiphenyl is a mammary carcinogen and was correctly predicted to be so by both the TSC-NC and TSC-NTSC models (Table 3). By chance it was also selected to be a non-target site carcinogen in the liver TSC-NTSC model where it was correctly predicted to be a carcinogen to other sites aside from the liver (Table 3).

4. Conclusions

The present study consisted of SAR analyses of organ-specific carcinogenesis based on two different selection methods for identifying inactive compounds for the model learning sets. The first selection routine generating the TSC-NC models analysed organ-specific carcinogens to non-carcinogens. The second method generating the TSC-NTSC models analysed carcinogens to one site carcinogens to all other sites but the target site. First, since both the TSC-NC and TSC-NTSC models mostly had concordance, sensitivity, and specific value ranging from the 0.70s to the 0.90s, it is evident that cat-SAR analyses of this type can lead to predictive models for both analyzing carcinogens and non-carcinogens, but also for organ-specific carcinogens.

The group of 12 TSC-NTSC models did not follow the typical SAR paradigm of comparing carcinogens to non-carcinogens but rather comparing two sets of carcinogens. Since both the active and inactive categories for these SAR models were populated with carcinogens, the use of 2-dimensional structural fragments was able to accurately distinguish between carcinogens to one site from those of other sites. Therefore, by contrasting chemicals that induced cancer to one site to those that induced cancer at other target sites, we speculate that the structural features related to the phenomena of mutagenicity (e.g., Ashby' structural alerts) are now accompanied by other features relating to targeting the carcinogen to a specific organ.

In general, the identified attributes derived from the TSC-NTSC models can be used to explore previously established or hypothesized mechanisms for chemically induced organ-specific cancer. These findings suggested that the TSC-NTSC models can identify structural attributes of chemicals that impart on them the ability to induce cancer at specific target sites, which are separable from those generally associated with carcinogenic potential (e.g., DNA-reactivity). For example, consider the premise of SAR studies that like structure begets like activity. The original question of “why do some carcinogens cause breast cancer?” is answered by the MC-NMC model wherein breast carcinogens have certain (and identifiable) structural moieties relating them specifically to breast carcinogenesis rather than to cancer at other sites. In other words, in this example, it is apparent that mammary carcinogens require different structural attributes; ones associated with “carcinogenicity” (e.g., DNA reactivity) and ones associated with “mammary gland” (e.g., estrogenicity). Moreover, the MC-NMC model presented here or other SAR models developed that contrast carcinogens to organ-specific ones (i.e., rather than carcinogens to non-carcinogens) have the potential to be useful tools to investigate organ-specificity. It is conceivable that toxicophore-defined sets of compounds that induce organ-specific carcinogenesis may mostly influence molecular target(s) found only in that specific organ (e.g., ligand-receptor or agent-adduct). As such, these congeneric chemicals therefore exhibit organ-specific activity because only that organ has the proper molecular target for chemical interaction. Models of this type may prove useful in anticancer drug development since they essentially contain chemical moieties that could facilitate targeting agents to a specific tumour site.

Finally, the cat-SAR expert system constructs models from a knowledge-based approach (i.e., knowledge contained in the learning set) rather than a hypothesis-driven approach. Thus, the derived models are not dependent upon previous knowledge or assumptions regarding a mechanism of action. However, it is worth mentioning that the chemicals selected for the modelling process were based on a hypothesis (i.e., that there are structural differences that relate specifically to mammary cancer that are different than those that relate to other types of cancer). Thus, in the case of the TSC-NTSC models, it was dependent upon a previous assumption (i.e., hypothesis) regarding a mechanism of action. In essence then, when considering further studies regarding the mechanisms of action of carcinogens, or in predictive analyses to assess whether or not a chemical has the potential to be a carcinogen to a specific organ, a tiered computational approach would be necessary that would first assess carcinogenicity and then assess organ specific carcinogenicity.

Acknowledgments

This research was supported by the National Institutes of Health (P20 RR018733) and the Department of Defense Breast Cancer Research Program (DAMD17-01-0376). Views and opinions of, and endorsements by the author(s) do not reflect those of the US Army or the Department of Defense.

References

1. Cimino M. Comparative overview of current international strategies and guidelines for genetic toxicology testing for regulatory purposes. *Environ Mol Mutagen*. 2006; 47:362–390. [PubMed: 16649190]
2. Gold, LS.; Slone, TH.; Manley, NB.; Backman, GM.; Garfinkle, GB.; Rohrbach, L.; Ames, BN. The Carcinogenic Potency Database. In: Gold, LS.; Zeiger, E., editors. *Handbook of Carcinogenic Potency and Genotoxicity Databases*. CRC Press; Boca Raton: 1997. p. 1-605.
3. Williams-Brown S, Singh GK. Epidemiology of cancer in the United States. *Semin Oncol Nurs*. 2005; 21:236–242. [PubMed: 16293511]
4. Cunningham AR, Rosenkranz HS, Zhang YP, Klopman G. Identification of “genotoxic” and “non-genotoxic” alerts for cancer in mice: The carcinogenic potency database. *Mutat Res*. 1998; 398:1–17. [PubMed: 9626960]

5. Cunningham AR, Rosenkranz HS, Klopman G. Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutat Res.* 1998; 405:9–28. [PubMed: 9729240]
6. Gold, LS. [last accessed 11/07/07] Summary of Carcinogenic Potency Database by target organ. 2007. <http://potency.berkeley.edu/pdfs/CPDBPathology.pdf>
7. Contrera JF, Kruhlak NL, Matthews EJ, Benz RD. Comparison of MC4PC and MDL-QSAR rodent carcinogenicity predictions and the enhancement of predictive performance by combining QSAR models. *Regul Toxicol Pharmacol.* 2007; 49:172–182. [PubMed: 17703860]
8. McKinney JD, Richard A, Waller C, Newman MC, Gerberick F. The practice of structure activity relationships (SAR) in toxicology. *Toxicol Sci.* 2000; 56:8–17. [PubMed: 10869449]
9. Richard AM. Commercial toxicology prediction systems: A regulatory perspective. *Toxicol Lett.* 1998; 102–103:611–616.
10. Richard AM. Application of artificial intelligence and computer-based methods to predicting chemical toxicity. *Knowl Eng Rev.* 1999; 14:307–317.
11. Benfenati E, Benigni R, Demarini DM, Helma C, Kirkland D, Martin TM, Mazzatorta P, Ouedraogo-Arras G, Richard AM, Schilter B, Schoonen WGEJ, Snyder RD, Yang C. Predictive models for carcinogenicity and mutagenicity: Frameworks, state-of-the-art, and perspectives. *J EnvironSci Health, Part C.* 2009; 27:57–90.
12. Benigni R, Netzeva TI, Benfenati E, Bossa C, Franke R, Helma C, Hulzebos E, Marchant C, Richard A, Woo Y-T, Yang C. The expanding role of predictive toxicology: An update on the (Q)SAR models for mutagens and carcinogens. *J Environ Sci Health C.* 2007; 25:53–97.
13. Piegorsch, WW.; Zeiger, E. Measuring intra-assay agreement for the Ames Salmonella assay. In: Hotham, L., editor. *Statistical Methods in Toxicology.* Springer-Verlag; Heidelberg: 1991. p. 35-41.
14. Gold LS, Wright C, Bernstein L, deVeciana M. Reproducibility of results in near-replicate carcinogenesis bioassay. *J Natl Cancer Inst.* 1987; 78:1149–1158. [PubMed: 3473255]
15. Gold, LS.; Slone, TH.; Ames, BN. Overview and update of analyses of the carcinogenic potency database. In: Gold, LS.; Zeiger, E., editors. *Handbook of Carcinogenic Potency and Genotoxicity Databases.* CRC Press; New York: 1997. p. 661-693.
16. Cabello G, Valenzuela M, Vilaxa A, Durán V, Rudolph I, Hrepic N, Calaf G. A rat mammary tumor model induced by the organophosphorous pesticides parathion and malathion, possibly through acetylcholinesterase inhibition. *Environ Health Perspect.* 2001; 109:471–479. [PubMed: 11401758]
17. Benigni R, Bossa C. Predictivity of QSAR. *J Chem Inf Model.* 2008; 48:971–980. [PubMed: 18426198]
18. Milan C, Schifanella O, Roncaglioni A, Benfenati E. Comparison and possible use of in silico tools for carcinogenicity within REACH legislation. *J Environ SciHealth, Part C.* 2011; 29:300–323.
19. Miller, JA.; Miller, EC. Ultimate chemical carcinogens as reactive mutagenic electrophiles. In: Hiatt, HH.; Watson, JD.; Winsten, JA., editors. *Origins of Human Cancer.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor; New York: 1977. p. 605-627.
20. Benigni R, Bossa C, Tcheremenskaia O, Giuliani A. Alternatives to the carcinogenicity bioassay: in silico methods, and the in vitro and in vivo mutagenicity assays. *Exp Op Drug Metab Toxicol.* 2010; 6:1–11.
21. Sasaki YF, Sekihashi K, Izumiyama F, Nishidate E, Saga A, Ishida K, Tsuda S. The comet assay with multiple mouse organs: comparison of comet assay results and carcinogenicity with 208 chemicals selected from the IARC monographs and U.S. NTP Carcinogenicity Database. *Crit Rev Toxicol.* 2002; 30:629–799. [PubMed: 11145306]
22. Sekihashi K, Yamamoto A, Matsumura Y, Ueno S, Watanabe-Akanuma M, Kassie F, Knasmüller S, Tsuda S, Sasaki YF. Comparative investigation of multiple organs of mice and rats in the comet assay. *Mutat Res.* 2002; 517:53–75. [PubMed: 12034309]
23. Hemminki, K.; Thilly, WG. Implications of results of molecular epidemiology on DNA adducts, their repair and mutations for mechanisms of human cancer. In: Buffler, P.; Rice, J.; Baan, R.; Bird, M.; Boffetta, P., editors. *Mechanisms of Carcinogenesis: Contributions of Molecular Epidemiology.* International Agency for Research on Cancer; Lyon: 2004. p. 217-235.
24. Swenberg, JA. Toxicological considerations in the application and interpretation of DNA adducts in epidemiological studies. In: Buffler, P.; Rice, J.; Baan, R.; Bird, M.; Boffetta, P., editors.

- Mechanisms of Carcinogenesis: Contributions of Molecular Epidemiology. International Agency for Research on Cancer; Lyon: 2004. p. 237-246.
25. Weiss G. Acrylamide in food: Uncharted territory. *Science*. 2002; 297:27. [PubMed: 12098675]
 26. Tareke E, Rydberg P, Karlsson P, Eriksson S, Tornqvist M. Acrylamide: A cooking carcinogen? *Chem Res Toxicol*. 2000; 13:517–522. [PubMed: 10858325]
 27. Fohgelberg P, Rosén J, Hellenäs KE, Abramsson-Zetterberg L. The acrylamide intake via some common baby food for children in Sweden during their first year of life--an improved method for analysis of acrylamide. *Food Chem Toxicol*. 2005; 43:951–959. [PubMed: 15811575]
 28. Lafferty JS, Kamendulis LM, Kaster J, Jiang J, Klaunig JE. Subchronic acrylamide treatment induces a tissue-specific increase in DNA synthesis in the rat. *Toxicol Lett*. 2004; 154:95–103. [PubMed: 15475183]
 29. Doerge DR, Gamboa da Costa G, McDaniel LP, Churchwell MI, Twaddle NC, Beland FA. DNA adducts derived from administration of acrylamide and glycidamide to mice and rats. *Mutat Res/Gen Toxicol Environ Mutag*. 2005; 580:131–141.
 30. Manière I, Godard T, Doerge DR, Churchwell MI, Guffroy M, Laurentie M, Poul J-M. DNA damage and DNA adduct formation in rat tissues following oral administration of acrylamide. *Mutat Res/Gen Toxicol Environ Mutag*. 2005; 580:119–129.
 31. Young JF, Tong W, Fang H, Xie Q, Pearce B, Hashemi R, Beger RD, Cheeseman MA, Chen JJ, Chang Y-cI, Kodel RL. Building an organ-specific carcinogenic database for SAR analyses. *J Toxicol Environ Health A*. 2004; 67:1363–1389. [PubMed: 15371237]
 32. Gold LS, Manley NB, Slone TH, Ward JM. Compendium of chemical carcinogens by target organ: Results of chronic bioassays in rats, mice, hamsters, dogs, and monkeys. *Toxicol Pathol*. 2001; 29:639–652. [PubMed: 11794380]
 33. Cunningham AR, Moss ST, Iype SA, Qian G, Qamar S, Cunningham SL. Structure-activity relationship analysis of rat mammary carcinogens. *Chem Res Toxicol*. 2008; 21:1970–1982. [PubMed: 18759503]
 34. Gold LS, Sawyer CB, Magaw R, Backman GM, deVeciana M, Levinson R, Hooper NK, Havender WR, Bernstein L, Peto R, Pike MC, Ames BN. A carcinogenic potency database of the standardized results of animal bioassays. *Environ Health Perspect*. 1984; 58:9–319. [PubMed: 6525996]

\$watermark-text

\$watermark-text

\$watermark-text

Table 1

LNO and cross validation results for the target site carcinogen non-carcinogen (TSC-NC) SAR models.

Model	LOO Best Concordance				LOO Balanced Sensitivity and Specificity				LMO			LNO		
	Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance
Clitoral Gland														
1 (Best)	0.84 (16/19)	0.95 (18/19)	0.90 (34/38)		0.84 (16/19)	0.84 (16/19)	0.84 (32/38)		0.84 (1.6/1.9)	0.83 (1.5/1.8)	0.83 (3.1/3.7)	1.00 (18/18)	1.00 (18/18)	1.00 (36/36)
2	0.95 (18/19)	0.72 (13/18)	0.84 (31/37)		0.79 (15/19)	0.85 (17/20)	0.82 (32/39)							
3	0.83 (15/18)	0.85 (17/20)	0.84 (32/38)		0.83 (15/18)	0.85 (17/20)	0.84 (32/38)							
Average (SD)	0.87 (0.06)	0.84 (0.11)	0.86 (0.03)		0.82 (0.03)	0.84 (0.01)	0.83 (0.01)							
Esophagus														
1 (Best)	0.91 (32/35)	1.00 (36/36)	0.96 (68/71)		0.94 (33/35)	0.94 (34/36)	0.94 (67/71)		0.91 (2.9/3.2)	0.92 (3.0/3.3)	0.91 (5.9/6.5)	0.94 (34/36)	0.94 (34/36)	0.94 (68/72)
2	0.89 (31/35)	1.00 (35/35)	0.94 (66/70)		0.91 (31/34)	0.97 (32/33)	0.94 (63/67)							
3	0.89 (31/35)	1.00 (33/33)	0.94 (64/68)		0.91 (32/35)	0.91 (30/33)	0.91 (62/68)							
Average	0.90 (0.02)	1.00 (0.00)	0.95 (0.01)		0.92 (0.02)	0.94 (0.02)	0.93 (0.02)							
Hematopoietic														
1	0.69 (36/52)	0.77 (41/53)	0.73 (77/105)		0.71 (37/52)	0.72 (38/53)	0.71 (75/105)							
2 (Best)	0.83 (40/48)	0.80 (41/51)	0.82 (81/99)		0.81 (39/48)	0.82 (42/51)	0.82 (81/99)		0.74 (3.2/4.4)	0.77 (3.5/4.6)	0.75 (6.7/9.0)	0.95 (42/44)	0.96 (49/51)	0.95 (91/95)
3	0.69 (34/49)	0.87 (46/53)	0.78 (80/102)		0.69 (34/49)	0.87 (46/53)	0.78 (80/102)							
Average	0.74 (0.08)	0.82 (0.05)	0.78 (0.04)		0.74 (0.06)	0.80 (0.08)	0.77 (0.05)							
Kidney														
1	0.65 (53/81)	0.72 (58/81)	0.69 (111/162)		0.67 (54/81)	0.67 (54/81)	0.67 (108/162)							
2 (Best)	0.60 (48/80)	0.78 (64/82)	0.69 (112/162)		0.68 (52/77)	0.68 (56/82)	0.68 (108/159)		0.66 (4.9/7.5)	0.67 (5.1/7.8)	0.66 (10.1/15.2)	1.00 (79/79)	1.00 (84/84)	1.00 (163/163)
3	0.68 (52/77)	0.62 (49/79)	0.65 (101/156)		0.62 (48/77)	0.67 (53/79)	0.65 (101/156)							
Average	0.64 (0.04)	0.71 (0.08)	0.67 (0.02)		0.66 (0.03)	0.67 (0.01)	0.66 (0.02)							
Large Intestine														
1	0.81 (22/27)	0.85 (23/27)	0.83 (45/54)		0.78 (21/27)	0.81 (21/26)	0.79 (42/53)							
2	0.93 (26/28)	0.81 (22/27)	0.83 (48/58)		0.81 (21/26)	0.85 (22/26)	0.83 (43/52)							
3 (Best)	0.89 (25/28)	0.83 (24/29)	0.86 (49/57)		0.85 (23/27)	0.84 (21/25)	0.85 (44/52)		0.81 (1.8/2.3)	0.83 (1.8/2.2)	0.81 (3.6/4.4)	1.00 (27/27)	1.00 (25/25)	1.00 (52/52)
Average	0.88 (0.06)	0.83 (0.02)	0.84 (0.02)		0.81 (0.04)	0.83 (0.02)	0.82 (0.03)							

\$watermark-text

\$watermark-text

\$watermark-text

Model	LOO Best Concordance				LOO Balanced Sensitivity and Specificity				LMO			LNO		
	Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance

Liver														
1	0.67 (126/189)	0.84 (159/190)	0.75 (285/379)		0.74 (141/191)	0.74 (141/191)	0.74 (282/382)							
2	0.76 (139/183)	0.77 (148/193)	0.76 (287/376)		0.76 (139/183)	0.76 (147/193)	0.76 (286/376)							
3 (Best)	0.71 (137/192)	0.81 (158/196)	0.76 (295/388)		0.75 (141/189)	0.75 (143/191)	0.75 (284/380)		0.72 (12.7/17.8)	0.75 (13.2/17.8)	0.73 (26.0/35.6)	0.97 (187/193)	0.97 (187/193)	0.97 (374/386)
Average	0.71 (0.05)	0.80 (0.04)	0.76 (0.01)		0.75 (0.01)	0.75 (0.01)	0.75 (0.01)							
Lung														
1	0.68 (34/50)	1.00 (50/50)	0.84 (84/100)		0.78 (40/51)	0.79 (38/48)	0.79 (78/99)							
2	0.87 (41/47)	0.81 (38/47)	0.84 (79/94)		0.85 (40/47)	0.81 (38/47)	0.83 (78/94)							
3	0.80 (40/50)	0.90 (46/51)	0.86 (86/100)		0.80 (40/50)	0.82 (41/50)	0.81 (81/100)		0.77 (3.6/4.5)	0.79 (3.6/4.6)	0.79 (7.2/9.1)	0.94 (46/49)	0.94 (46/49)	0.94 (92/98)
Average	0.78 (0.10)	0.90 (0.10)	0.85 (0.01)		0.81 (0.03)	0.81 (0.01)	0.81 (0.02)							
Mammary														
1 (Best)	0.75 (77/102)	0.99 (89/90)	0.83 (166/201)		0.79 (77/97)	0.83 (81/98)	0.81 (158/195)		0.78 (6.8/8.7)	0.69 (6.4/9.1)	0.74 (13.1/17.7)	0.78 (70/89)	0.84 (83/98)	0.81 (153/187)
2	0.66 (66/100)	0.92 (94/102)	0.79 (160/202)		0.76 (78/103)	0.79 (79/100)	0.77 (157/203)							
3)	0.63 (64/102)	0.96 (99/103)	0.80 (163/205)		0.75 (71/95)	0.76 (71/93)	0.76 (142/188)							
Average	0.68 (0.07)	0.96 (0.03)	0.80 (0.02)		0.77 (0.02)	0.79 (0.03)	0.78 (0.03)							
Nasal Cavity														
1	0.79 (38/48)	0.92 (44/48)	0.85 (82/96)		0.80 (37/46)	0.80 (36/45)	0.80 (73/91)							
2	0.64 (27/42)	1.00 (46/46)	0.83 (73/88)		0.79 (38/48)	0.83 (39/47)	0.81 (77/95)							
3 (Best)	0.75 (36/48)	0.94 (45/48)	0.84 (81/96)		0.81 (39/48)	0.81 (39/48)	0.81 (78/96)		0.78 (3.4/4.3)	0.75 (3.3/4.3)	0.77 (6.63/8.6)	0.88 (42/48)	0.88 (42/48)	0.88 (84/96)
Average	0.73 (0.08)	0.95 (0.04)	0.84 (0.01)		0.80 (0.01)	0.81 (0.01)	0.81 (0.01)							
Nervous System														
1	0.87 (13/15)	0.79 (15/19)	0.82 (28/34)		0.76 (16/21)	0.76 (16/21)	0.76 (32/42)							
2 (Best)	0.83 (15/18)	0.90 (18/20)	0.87 (33/38)		0.83 (15/18)	0.85 (17/20)	0.84 (32/38)		0.75 (1.3/1.8)	0.85 (1.5/1.9)	0.78 (2.82/3.62)	0.90 (18/20)	0.91 (19/21)	0.90 (37/41)
3	0.67 (12/18)	0.81 (17/21)	0.74 (29/39)		0.67 (14/21)	0.67 (14/21)	0.67 (28/42)							
Average	0.79 (0.11)	0.83 (0.06)	0.81 (0.06)		0.75 (0.08)	0.76 (0.09)	0.76 (0.09)							
Small Intestine														
1	0.72 (18/25)	0.93 (25/27)	0.83 (43/52)		0.81 (21/26)	0.83 (24/29)	0.82 (45/55)							
2 (Best)	0.74 (20/27)	1.00 (26/26)	0.87 (46/53)		0.82 (23/28)	0.83 (24/29)	0.83 (47/57)		0.72 (1.6/2.3)	0.77 (1.9/2.4)	0.73 (3.5/4.8)	0.83 (24/29)	0.83 (24/29)	0.83 (48/58)

Model	LOO Best Concordance				LOO Balanced Sensitivity and Specificity				LMO		LNO	
	Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance	
3	0.88 (23/26)	0.78 (21/27)	0.83 (44/53)		0.81 (21/26)	0.81 (22/27)	0.81 (43/53)					
Average	0.78 (0.09)	0.90 (0.11)	0.84 (0.02)		0.81 (0.01)	0.82 (0.01)	0.82 (0.01)					
Uterus												
1 (Best)	0.83 (19/23)	0.83 (19/23)	0.83 (38/46)		0.83 (19/23)	0.83 (19/23)	0.83 (38/46)					
2	0.91 (19/21)	0.79 (19/24)	0.84 (38/45)		0.80 (20/25)	0.76 (19/25)	0.78 (39/50)					
3	0.88 (22/25)	0.65 (15/23)	0.77 (37/48)		0.76 (19/25)	0.79 (20/25)	0.78 (39/50)		0.81 (1.7/2.2)	0.78 (1.7/2.2)	0.96 (23/24)	0.96 (44/46)
Average	0.87 (0.04)	0.76 (0.09)	0.81 (0.03)		0.80 (0.04)	0.79 (0.04)	0.80 (0.03)					
Averages												
1			0.79 (1061/1340)				0.77 (1028/1339)					
2			0.80 (1054/1322)				0.78 (1043/1329)					
3			0.79 (1061/1350)				0.77 (1014/1325)					
Best			0.84 (1092/1305)				0.79 (1050/1331)					

Numbers in the parentheses represent the number of correct predictions over the number of total predictions, for active, inactive and total compounds. For LMO, the numbers are represented as non-integers and are the average number of correct and total predictions over 1000 random 10% test sets.

\$watermark-text

\$watermark-text

\$watermark-text

Table 2
LNO and cross validation results for the target site carcinogen non-target site carcinogen (TSC-NTSC) SAR models.

Model	LOO Best Concordance				LOO Balanced Sensitivity and Specificity				LMO			LNO		
	Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance
Clitoral Gland														
1	0.80 (16/20)	0.85 (17/20)	0.83 (33/40)		0.80 (16/20)	0.82 (14/17)	0.81 (30/37)		0.77 (1.4/1.8)	0.86 (1.4/1.7)	0.81 (2.9/3.5)	0.89 (16/18)	0.90 (17/19)	0.89 (33/37)
2	0.95 (19/20)	0.81 (13/16)	0.89 (32/36)		0.83 (15/18)	0.89 (16/18)	0.86 (31/36)							
3	0.80 (16/20)	0.89 (17/19)	0.85 (33/39)		0.80 (16/20)	0.84 (16/19)	0.82 (32/39)							
Average	0.85 (0.09)	0.85 (0.04)	0.85 (0.03)		0.81 (0.02)	0.85 (0.03)	0.83 (0.03)							
Esophagus														
1	0.83 (30/36)	0.86 (31/36)	0.85 (61/72)		0.83 (30/36)	0.83 (30/36)	0.83 (60/72)							
2	0.88 (30/34)	0.85 (28/33)	0.87 (58/67)		0.86 (31/36)	0.86 (31/36)	0.86 (62/72)							
3	0.89 (31/35)	0.88 (29/33)	0.88 (60/68)		0.89 (31/35)	0.88 (29/33)	0.88 (60/68)		0.88 (2.8/3.2)	0.85 (2.5/3.0)	0.86 (5.4/6.2)	0.94 (32/34)	0.91 (31/34)	0.93 (63/68)
Average	0.87 (0.03)	0.86 (0.02)	0.87 (0.02)		0.86 (0.03)	0.86 (0.02)	0.86 (0.02)							
Hematopoietic														
1	0.72 (34/47)	0.75 (36/48)	0.74 (70/95)		0.70 (35/50)	0.75 (36/48)	0.72 (71/98)		0.66 (2.9/4.5)	0.70 (3.0/4.4)	0.69 (5.9/8.9)	0.82 (42/51)	0.80 (41/51)	0.81 (83/102)
2	0.80 (32/40)	0.64 (32/50)	0.64 (64/100)		0.64 (32/50)	0.64 (32/50)	0.64 (64/100)							
3	0.63 (33/52)	0.70 37/53)	0.67 (70/105)		0.63 (33/52)	0.66 (35/53)	0.65 (68/105)							
Average	0.72 (0.08)	0.70 (0.06)	0.68 (0.05)		0.66 (0.04)	0.68 (0.06)	0.67 (0.05)							
Kidney														
1	0.68 (56/82)	0.65 (54/83)	0.67 (110/165)		0.66 (53/80)	0.67 (50/75)	0.66 (103/155)							
2	0.67 (54/81)	0.65 (52/80)	0.66 (106/161)		0.65 (53/81)	0.65 (52/80)	0.65 (105/161)							
3	0.61 (47/77)	0.76 (61/80)	0.69 (108/157)		0.68 (56/82)	0.67 (56/84)	0.67 (112/166)		0.64 (5.0/7.9)	0.67 (5.2/7.9)	0.64 (10.2/15.8)	0.92 (78/85)	0.91 (77/85)	0.91 (155/170)
Average	0.65 (0.04)	0.69 (0.06)	0.67 (0.02)		0.67 (0.01)	0.66 (0.01)	0.66 (0.01)							
Large Intestine														
1	0.61 (17/28)	0.85 (23/27)	0.73 (40/55)		0.70 (19/27)	0.69 (20/29)	0.70 (39/56)							
2	0.79 (22/28)	0.62 (18/29)	0.70 (40/57)		0.65 (17/26)	0.68 (19/28)	0.67 (36/54)							
3	0.70 (19/27)	0.76 (19/25)	0.73 (38/52)		0.70 (19/27)	0.72 (18/25)	0.71 (37/52)		0.66 (1.4/2.2)	0.70 (1.47/2.2)	0.67 (2.9/4.4)	1.00 (27/27)	1.00 (26/26)	1.00 (53/53)
Average	0.70 (0.09)	0.74 (0.12)	0.72 (0.02)		0.69 (0.03)	0.70 (0.02)	0.69 (0.02)							

Model	LOO Best Concordance			LOO Balanced Sensitivity and Specificity			LMO			LNO		
	Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance
Liver												
1	0.76 (139/183)	0.70 (124/176)	0.73 (263/359)	0.72 (131/183)	0.73 (129/176)	0.72 (260/359)						
2	0.68 (120/177)	0.70 (129/184)	0.69 (249/361)	0.69 (121/176)	0.68 (125/184)	0.68 (246/360)						
3	0.76 (143/189)	0.69 (132/191)	0.72 (275/380)	0.72 (144/199)	0.71 (140/198)	0.72 (284/397)	0.70 (13.1/18.6)	0.69 (12.7/18.6)	0.70 (25.9/37.2)	0.78 (155/199)	0.80 (159/199)	0.79 (314/398)
Average	0.73 (0.05)	0.70 (0.01)	0.72 (0.02)	0.71 (0.02)	0.71 (0.03)	0.71 (0.02)						
Lung												
1	0.69 (35/51)	0.79 (41/52)	0.75 (76/102)	0.76 (37/49)	0.69 (33/48)	0.72 (70/97)						
2	0.72 (36/50)	0.86 (43/50)	0.79 (79/100)	0.78 (39/50)	0.78 (39/50)	0.78 (78/100)	0.74 (3.5/4.6)	0.71 (3.2/4.5)	0.72 (6.6/9.2)	0.80 (41/51)	0.84 (42/50)	0.82 (83/101)
3	0.60 (31/52)	0.76 (38/50)	0.68 (69/102)	0.68 (32/47)	0.68 (32/47)	0.68 (64/94)						
Average	0.67 (0.06)	0.80 (0.05)	0.74 (0.06)	0.74 (0.05)	0.72 (0.06)	0.73 (0.05)						
Mammary												
1	0.75 (77/103)	0.83 (85/103)	0.79 (162/206)	0.78 (77/99)	0.78 (79/101)	0.78 (156/200)	0.74 (6.9/9.3)	0.74 (7.0/10.0)	0.74 (13.9/18.9)	1.00 (100/100)	0.98 (98/100)	0.99 (198/200)
2	0.70 (66/94)	0.85 (80/94)	0.78 (146/188)	0.77 (77/100)	0.76 (75/99)	0.76 (152/199)						
3	0.70 (66/94)	0.85 (80/94)	0.78 (146/188)	0.77 (77/100)	0.76 (75/99)	0.76 (152/199)						
Average	0.72 (0.03)	0.84 (0.01)	0.78 (0.01)	0.77 (0.00)	0.77 (0.01)	0.77 (0.01)						
Nasal Cavity												
1	0.81 (39/48)	0.77 (37/48)	0.79 (76/96)	0.77 (37/48)	0.77 (37/48)	0.77 (74/96)	0.74 (3.0/4.1)	0.72 (3.2/4.4)	0.74 (6.2/8.5)	0.71 (29/41)	0.75 (36/48)	0.73 (65/89)
2	0.79 (34/43)	0.77 (36/47)	0.78 (70/90)	0.77 (33/43)	0.77 (36/47)	0.77 (69/90)						
3	0.71 (32/45)	0.86 (37/43)	0.78 (69/88)	0.73 (33/45)	0.77 (33/43)	0.75 (66/88)						
Average	0.77 (0.05)	0.80 (0.05)	0.78 (0.01)	0.76 (0.02)	0.77 (0.00)	0.76 (0.01)						
Nervous System												
1	0.76 (16/21)	0.62 (13/21)	0.69 (29/42)	0.67 (14/21)	0.68 (13/19)	0.68 (27/40)						
2	0.89 (17/19)	0.74 (14/19)	0.82 (31/38)	0.79 (15/19)	0.79 (15/19)	0.79 (30/38)	0.82 (1.5/1.7)	0.67 (1.2/1.7)	0.75 (2.6/3.4)	0.81 (13/16)	0.84 (16/19)	0.83 (29/35)
3	0.67 (14/21)	0.80 (16/20)	0.73 (30/41)	0.71 (15/21)	0.67 (14/21)	0.69 (29/42)						
Average	0.77 (0.11)	0.72 (0.09)	0.75 (0.06)	0.72 (0.06)	0.71 (0.07)	0.72 (0.06)						
Small Intestine												
1	0.76 (22/29)	0.89 (25/28)	0.82 (47/57)	0.81 (22/27)	0.81 (21/26)	0.81 (43/53)						
2	0.61 (17/28)	0.75 (21/28)	0.68 (38/56)	0.62 (18/29)	0.68 (19/28)	0.65 (37/57)						

Model	LOO Best Concordance				LOO Balanced Sensitivity and Specificity				LMO		LNO	
	Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Concordance		Sensitivity	Specificity	Sensitivity	Concordance
3	0.88 (23/26)	0.89 (25/28)	0.89 (48/54)		0.88 (23/26)	0.89 (25/28)	0.89 (48/54)		0.89 (2.0/2.2)	0.87 (2.0/2.3)	1.00 (27/27)	1.00 (55/55)
Average	0.75 (0.14)	0.85 (0.08)	0.80 (0.11)		0.77 (0.14)	0.79 (0.11)	0.78 (0.12)					
Uterus												
1	1.00 (22/22)	0.83 (19/23)	0.91 (41/45)		0.88 (21/24)	0.83 (20/24)	0.85 (41/48)		0.81 (1.9/2.3)	0.83 (1.9/2.3)	0.92 (22/24)	0.92 (43/47)
2	0.74 (17/23)	0.82 (18/22)	0.78 (35/45)		0.74 (17/23)	0.77 (17/22)	0.76 (34/45)					
3	0.60 (15/25)	0.68 (17/25)	0.64 (32/50)		0.60 (15/25)	0.68 (17/25)	0.64 (32/50)					
Average	0.78 (0.20)	0.77 (0.08)	0.78 (0.14)		0.72 (0.12)	0.78 (0.10)	0.75 (0.11)					
Averages												
1			0.76 (1008/1334)				0.74 (974/1311)					
2			0.73 (948/1299)				0.72 (944/1312)					
3			0.74 (978/1324)				0.73 (984/1354)					
Best			0.77 (1010/1309)				0.76 (1022/1353)					

Numbers in the parentheses represent the number of correct predictions over the number of total predictions, for active, inactive and total compounds. For LMO, the numbers are represented as non-integers and are the average number of correct and total predictions over 1000 random 10% test sets.