Published in final edited form as:

SAR QSAR Environ Res. 2014 ; 25(6): 489–506. doi:10.1080/1062936X.2014.898694.

Structure-Activity Relationship Models for Rat Carcinogenesis and Assessing the Role Mutagens Play in Model Predictivity

C. Alex Carrasquer¹, Kaylind Batey¹, Shahid Qamar¹, Albert R. Cunningham^{1,2,3,*}, and Suzanne L. Cunningham⁴

¹James Graham Brown Cancer Center, University of Louisville

²Department of Medicine, University of Louisville

³Department of Pharmacology and Toxicology, University of Louisville

⁴Gnarus Systems, Inc. Louisville, KY, 40202, USA

Abstract

We previously demonstrated that fragment based cat-SAR carcinogenesis models consisting solely of mutagenic or non-mutagenic carcinogens varied greatly in terms of their predictive accuracy. This led us to investigate how well the rat cancer cat-SAR model predicted mutagens and non-mutagens in their learning set. Four rat cancer cat-SAR models were developed: Complete Rat, Transgender Rat, Male Rat, and Female Rat, with leave-one-out (LOO) validation concordance values of 69%, 74%, 67%, and 73%, respectively. The mutagenic carcinogens produced concordance values in the range of 69–76% as compared to only 47–53% for non-mutagenic carcinogens. As a surrogate for mutagenicity comparisons between single site and multiple site carcinogen SAR models was analyzed. The LOO concordance values for models consisting of 1-site, 2-site, and 4+-site carcinogens were 66%, 71%, and 79%, respectively. As expected, the proportion of mutagens to non-mutagens also increased, rising from 54% for 1-site to 80% for 4+-site carcinogens. This study demonstrates that mutagenic chemicals, in both SAR learning sets and test sets, are influential in assessing model accuracy. This suggests that SAR models for carcinogens may require a two-step process in which mutagenicity is first determined before carcinogenicity can be accurately predicted.

Keywords

Structure-Activity Relationship; Cancer; Mutagenic Carcinogens; Non-Mutagenic Carcinogens; Multisite Carcinogens

1. Introduction

Identifying potential human carcinogens is a difficult task. Although data from epidemiological studies have determined the carcinogenic effects of certain environmental compounds, the data is limited and retrospective. As estimated by the Toxic Substances

*Corresponding author: James Graham Brown Cancer Center, 505 South Hancock Street, University of Louisville, Louisville, KY 40202, al.cunningham@louisville.edu, voice: 502-852-7511, fax: 502-852-7979.

Control Act inventory, there are approximately 84,000 synthesized or characterized compounds in industrial and commercial use in the United States [1, 2]. Of these, about 2,000 have been tested in standardized two year carcinogenicity studies [3]. The European Union has also created a plan to develop new policies to limit the level of human health hazards. The new REACH (Regulation, Evaluation, Authorization, and Restrictions of Chemicals) program was initiated in 2007 mainly due to the tens of thousands of industrial compounds that may have unknown adverse health effects [4, 5].

For the last few decades, the two year rodent bioassay for carcinogen determination has been the “gold standard” for assessing the potential of chemicals to be human carcinogens. Rodents provide a genetically similar model in which a compound’s mechanism of action can be studied and characterized in humans [6]. In many cases, the first evidence of a human toxin’s adverse effects came from experimental animal studies. In fact, all known human carcinogens that have been tested adequately in laboratory rodents are also carcinogenic in those experimental animals [7]. On the other hand however, all compounds shown to induce cancer in laboratory animals are not necessarily human carcinogens [8, 9]. For all of the advantages associated with animal testing, there are also disadvantages. Animal two year bioassays are time consuming, extremely costly, have low throughput, and use large numbers of animals [3].

As outlined by the National Research Council (NRC) in its 2007 report *Toxicity Testing in the 21st Century: A Vision and a Strategy*, to improve the efficiency of assessing potential human hazards of environmental chemicals, it is necessary to develop alternative *in vitro* methods to reduce and eventually eliminate *in vivo* rodent bioassays [10]. The overall vision involves five components: chemical characterization, toxicity testing, dose-response and extrapolation modeling, human exposure data, and risk contexts. As described in the report, chemical characterization is the process by which data is collected on physical and chemical properties, including structural attributes, environmental concentrations, toxic potentials, and mechanistic action of toxicity for the purpose of identifying untested compounds with similar characteristics [10].

Various computational techniques, such as quantitative and qualitative structure activity relationship (SAR) modeling, are ideal methods to realize this goal [11–13]. Generally, SAR modeling identifies a link between chemical structural characteristics and biological activity of congeneric and noncongeneric compounds. Quantitative SAR models use physicochemical properties to form mathematical relationships that are analyzed by regression equations to give a more precise biological activity value. Qualitative SAR generally identifies the difference of compounds in two categories (active or inactive) to predict “yes” or “no” as to whether a compound has toxic potential. Once the structurally relevant features related to that learning set are identified, they can be used to predict the activity of untested chemicals and study toxic mechanisms. Overall, SAR models can be used to 1) to predict a compound’s toxic potential and to understand mechanisms of action; and 2) to screen compounds in a short time period and cost effective manner [11–13]. Therefore, SAR modeling has become internationally recognized as a valid method for identifying and linking untested compounds with a distinct toxic endpoint.

The development and application of SAR expert systems for the purpose of predictive toxicology and screening has become more widespread and accepted recently. Over 25 years ago, the Computer Automated Structure Evaluation (CASE) program was developed by Rosenkranz and Klopman [14, 15]. The CASE program successfully created models by using categorized active and inactive chemicals in a learning set to identify 2-dimensional structural features called biophores that were associated with a particular biological, pharmacological, or toxicological activity. These biophores were the key to correctly predicting unknown compounds as potential toxins. Other expert systems were also developed including TOPKAT [16], DEREK [17], along with human expert systems such as John Ashby's "structural alerts" (SA), which identify functional groups as potential markers for genotoxic carcinogenesis [18–20]. A recent study has also demonstrated that incorporating high-throughput screening (HTS) data with QSAR analysis of rodent carcinogens can improve the overall predictivity of the model [21]. Furthermore, van Leeuwen et al. As SAR expert systems are gaining more acceptance in the toxicological field, they are increasingly being used by regulatory agencies worldwide for a variety of purposes including cancer studies at the National Toxicology Program (NTP) [22], preclinical toxicological studies at the Food and Drug Administration (FDA) [23, 24], and for the identification of potential environmental toxins at the Environmental Protection Agency (EPA) [25].

We previously reported SAR models for rats and mice derived from CASE/MultiCASE analyses of the first five plots of the Carcinogenic Potency Database (CPDB) [26–33]. The CPDB is an internationally recognized public resource that has analyzed and consolidated 6540 long-term animal cancer bioassays performed on 1547 different chemicals from general literature as well as by the NCI/NTP. These models provided unique insight into the structural underpinnings of rodent carcinogenesis. The approach that we have taken in developing the cat-SAR program diverges from some existing commercial SAR expert systems that use defined and proprietary modeling technology since descriptor selection and mathematical model derivation are transparent and controllable [34–36]. Thus, with cat-SAR, the control and selection of modeling parameters facilitate the ability to explore rigorously the relationships between chemical structure and biological activity. Ultimately, this rationale negates any *a priori* requirements that a given set of data must fit the attributes of a predefined and often proprietary modeling process.

The cat-SAR models are built through a comparison of structural features found among categorized compounds in the model's learning set. Generically, these categories are biologically *active* and *inactive* compounds. In this study, we examined the differences in structural attributes of chemicals determined or classified to be carcinogens and noncarcinogens in rats. All data used for this study were obtained from the DssTox website [37] and all the information about the compounds has been compiled in the CPDB [27–33]. Previously, in an early human expert SAR study, Ashby *et al.* analyzed chemicals in both the NCI/NTP and CPDB datasets for correlations between structural characteristics and their carcinogenicity and mutagenicity. It was determined that there was a 90% correlation between chemical structure and *Salmonella* mutagenicity (genetic toxicity) among all 222 chemicals [19, 20]. Later, Ashby *et al.* extended their analysis to 552 compounds and found similar conclusions in that genotoxicity is very closely linked to carcinogenesis [19].

As described by Zeiger the current battery of tests for assessing carcinogenic potential include the *Salmonella* test with or without *E. coli*, the *in vitro* mammalian cell mouse lymphoma assay and/or the *in vitro* mammalian cell chromosome assay, and the *in vivo* rodent bone marrow chromosome aberration or micronucleus assay. These tests account for mutagenic carcinogens but do not consider non-mutagenic chemicals which also have carcinogenic potential by indirectly affecting DNA [38]. In contrast to mutagenic carcinogens, it is quite difficult to use SAR modeling to predict non-mutagenic carcinogens, which cause cancer through multiple non-DNA damaging mechanisms. Combes included cytotoxic activity along with the inhibition of gap-junctional intercellular communication, inhibition of tubulin polymerization, modulation of apoptosis and induction of cell proliferation, and binding to specific molecular targets (receptors). Some of the non-mutagenic mechanisms of carcinogenesis include protein kinase C, the estrogen, peroxisome-proliferator and tubulin protein receptors [39]. Due to various mechanisms through which a compound would bind to a non-DNA target, there are no small fragments or “structural alerts” associated with non-genotoxic carcinogenicity as there are with carcinogenicity. Benigni notes that the *Salmonella* assay is the most predictive assay for DNA-reactive carcinogens and that no other mutagenicity based test exists for non-DNA-reactive compounds [40], while Zeiger agrees by stating that additional test besides *Salmonella* do not help detecting *Salmonella*-negative carcinogens from *Salmonella*-negative non-carcinogens [38]. Finally, Woo pointed out that the identification of structural attributes of non-DNA reactive carcinogens (which can be used for prediction) is not as advanced as that for genotoxic carcinogens [41].

However, in a recent study we investigated a set of compounds that had been tested for both rat carcinogenicity and *Salmonella* mutagenicity. Using a novel SAR modeling method in which ligand-receptor interactions were used as descriptors in the modeling process, we demonstrated that this novel receptor-based method was able to predict non-mutagenic carcinogens with a higher efficiency than the traditional fragment model. In fact, the ligand model produced a concordance value of 71% as compared to 55% for the fragment model [42].

The current study was developed to further analyze the effects of mutagenicity on rat carcinogenicity in SAR modeling. This study includes analysis of all rat carcinogens, as well as an investigation on how dividing compounds into separate gender models affects the outcome. Furthermore, the relationship of mutagenicity and rat cancer was investigated through analysis of how well the mutagenic and non-mutagenic carcinogens were predicted during the SAR modeling validation process. Finally, a relationship between the number of sites at which a compound produced cancer and both mutagenicity and the concordance values for the cat-SAR models was determined.

2. Methods

2.1 Rat Carcinogenesis Learning Sets

All compounds for each of the various models’ learning sets were obtained from the DSSTox CPDB [37]. The CPDB is a collection of chemical carcinogenesis data obtained from bioassays on several species performed by the NTP as well as an ongoing search and

compilation of appropriate literature data for chemical carcinogenicity [43]. For models developed here, compounds in the CPDB designated as carcinogens had to be positive in either male or female rats and non-carcinogens had to be either negative in both genders or negative in one gender and not tested in the other. To get a more comprehensive understanding of the carcinogenic effect of chemicals on rats, three different rat learning sets were created. The first was an overall rat learning set that took into consideration all compounds that were carcinogenic in both male and females. The second and third models were gender specific and only considered chemicals that were tested in either male or female rats. Furthermore, the relationship between mutagenicity and carcinogenicity was investigated by creating three tumor site-number dependent models, in which characteristics of compounds that caused cancer at one site were compared to compounds that caused cancer at two sites and four or more sites.

The cat-SAR learning sets consisted of the chemical name, its structure as a .MOL2 file, and its categorical designation (*e.g.*, carcinogen or non-carcinogen). 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 Rat Model Learning Sets

Rat Model 1—The Complete Rat model consisted of 924 compounds that were tested for carcinogenicity in male and/or female rats with 531 carcinogens and 393 non-carcinogens. Compounds designated as carcinogens had to be positive in either male or female rats and non-carcinogens had to be either negative in both genders or negative in one gender and not tested in the other.

Rat Model 2—The Transgender Rat model was composed of compounds with consistent activity between both male and female rats. Therefore, compounds that were designated as carcinogens had to be positive in both genders or at least positive in one gender and not tested in the other. The noncarcinogens had to be either negative in both genders or at least negative in one gender and not tested in the other. In the end, 97 compounds that were observed to have divergent activity between genders in the Complete Rat model were deleted, leaving a total of 827 compounds in this model with 434 carcinogens and 393 non-carcinogens.

2.3 Gender Specific Rat Model Learning Sets

Male Rat Model—The Male Rat model consisted of 857 compounds that were tested for carcinogenicity in male rats with 433 carcinogens and 424 non-carcinogens. Compounds designated as carcinogens had to be positive and non-carcinogens had to be negative for cancer in male rats.

Female Rat Model—The Female Rat model consisted of 836 compounds that were tested for carcinogenicity in female rats with 377 carcinogens and 459 non-carcinogens. Compounds designated as carcinogens had to be positive and non-carcinogens had to be negative for cancer in female rats.

2.4 Tumor Site Models

1-Site Model—The 1-site tumor model consisted of 534 compounds that were tested for carcinogenicity in rats where 267 compounds were carcinogens at 1 site and 267 compounds were non-carcinogens that were randomly selected, creating an equal number of active and inactive compounds.

2-Site Model—The 2-site tumor model consisted of 196 compounds that were tested for carcinogenicity in rats where 98 compounds were carcinogens at 2 sites and 98 compounds were non-carcinogens that were randomly selected, creating an equal number of active and inactive compounds.

4+-Site Model—The 4+-site tumor model consisted of 174 compounds that were tested for carcinogenicity in rats where 87 compounds were carcinogens at four or more sites and 87 compounds were non-carcinogens that were randomly selected, creating an equal number of active and inactive compounds.

2.5 In Silico Chemical Fragmentation and the Compound-Fragment Data Matrix

Using the Tripos Sybyl HQSAR module, each chemical in a model's learning set 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, atomic connections, and explicate hydrogen atoms.

Upon completion of the fragmentation routine, a Sybyl HQSAR add-on produced a compound-fragment data matrix as a text file. In the matrix, the rows are intact chemicals and the columns represent 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 analyzed, 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.

2.6 Rule Optimization for 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 (Number Rule) is the number of compounds in the learning set in which a fragment is found. For this study, it was set between one and eight compounds. We reasoned that by considering fragments that came from between one and eight compounds in the learning set,

models derived in the one to four range would be more inclusive (i.e., higher coverage) while those in the five to eight range would be more accurate (i.e., higher concordance). We note that in previous cat-SAR analyses the fragment number was arbitrarily set to three.

The second rule (Proportion 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% with increasing intervals of 5%. 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 might 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).

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. Thus to best classify compounds back to an active or inactive category, we identify an optimal cut-off point that best separates the probabilistic prediction of active and inactive compounds based on leave-one-out (LOO) validation data.

In the study the Rule Optimization method created 800 models ranging from 1–50%–50% to 8–95%–95% (Number Rule-Active Proportion Rule-Inactive Proportion Rule). Due to the diverse characteristics among compounds in the learning sets and unbalanced number of Active and Inactive compounds for the different endpoints, the number of “important” active and inactive fragments can vary greatly. Choosing a standard set of selection rules or descriptor parameters for every model (e.g., 3–90%–90%) could limit the cat-SAR user’s ability to identify the most concordant model for that endpoint and thus decrease the predictive ability of that learning set.

The Rule Optimization method allows the cat-SAR user to identify an appropriate model for study based on an analysis in which an defined amount of data comes from Active and Inactive compounds alike, thus developing a predictive model with high coverage. The optimal model for each learning set was determined by LOO validation considering the following: 1) the model produced the highest concordance value in which the difference between the sensitivity and specificity values was close to zero; 2) the sensitivity, specificity, and concordance values were greater than or equal to 60%; and 3) the number of total predicted compounds (Coverage) was greater than 90% of the number compounds in the learning set.

It should be noted that by developing and analyzing a large number of potential models, the possibility exists of over-fitting the data. This may account for the variation in validation statics between the LNO and the LOO and LMO analyses (see Table 1). However, this difference between LNO and leave-10%-out concordance values in previous studies was also

observed. For example in MultCASE analyses of rodent carcinogens based on a single model, the difference between LNO to leave-10%-out concordance values was 95% to 64% (rats) and ~95% to ~70% for mice [26, 44]. Hence whether or not the cat-SAR models described herein are over-fit model or influenced by other phenomena of the modeling process remains to be determined.

2.7 Model Validation

Four validation routines were conducted for each model: a self-validating, leave-none-out (LNO); two cross-validation routines, LOO and leave-many-out (LMO); and an external validation (EXT). For the LNO routine, each chemical is predicted one at a time by a model based on the entire dataset, including the compound itself. 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 predicted using the n-1 model. Predicted vs. experimental values for each chemical were compared and the model's concordance, sensitivity, and specificity were determined.

Based on the Rule Optimization and LOO data, final selection rules were selected and final models were assessed with LMO cross validation. For LMO cross-validation, randomly selected sets of 10% of the chemicals were removed from the total fragment set, and the n-10% model was derived. Again, the activity of each of the removed chemicals was predicted using the n-10% model. Therefore, 90% of the compounds were used as the training set to predict the activities of the test set, which were the remaining 10% of the compounds. This process was repeated 1000 times. Predicted vs. experimental values for the chemicals in the test sets were compared and the concordances, sensitivities, and specificities of all 1000 models were determined and averaged to produce a final value.

For the EXT validation test, a random set of 10% of chemicals in the learning sets was removed (Test Set) with the remaining 90% of the compounds used to develop a model (Training Set) and predict the activity of those left out. A true external validation was created by taking the average of sensitivity, specificity, and concordance values of 10 random sets. In contrast to the LMO, none of the compounds and thus no fragment information from the Test Set were used by the Training Set to develop a model.

3. Results and Discussion

The present study was conducted to determine the applicability of the cat-SAR expert system to the task of analyzing and identifying rat chemical carcinogens and the role that mutagenic and non-mutagenic chemicals play in the determined accuracy of developed SAR models. The compounds collected for this study were obtained from the CPDB [27–33], which consists of 924 compounds that have been tested in rats, of which 857 and 836 compounds have been tested in male and female rats, respectively.

3.1 Rat Model

The LNO internal validation for the Complete Rat model produced a concordance between experimental and predictive values of 86% with a sensitivity of 86% and specificity of 86%. Evaluation of the LOO validations demonstrated that the model had a concordance of 69%

with a sensitivity of 70% and specificity of 68%. The LOO model made correct predictions on 570 compounds out of a total of 827 predictions, with a coverage of 90% (827 predicted compounds out of a possible 924) and the LMO produced a concordance of 68%.

The Transgender Rat model was created by removing compounds that produced positive cancer results in one gender and negative cancer results in the other. Therefore, the model's learning set consisted of 434 active compounds and 393 inactive compounds. The LNO model produced a concordance of 90% with a sensitivity of 90% and a specificity of 89%. The LOO of this model produced a concordance of 74% with a sensitivity of 73% and a specificity of 74% (Table 1) which were significantly higher than the Complete Rat model ($p < 0.05$).

3.2 Male and Female Rat Model

Analysis of the LNO validations for the Male Rat models demonstrated that the model produced a concordance of 79% with a sensitivity of 83% and specificity of 76% (Table 1). The LOO for the Male Rat model produced a concordance of 67% with a sensitivity of 66% and a specificity of 67% (Table 1) and the LMO for the Male Rat model produced a concordance of 67%. Analysis of the LNO validations for the Female Rat model demonstrated that the Rule Optimized model produced a concordance of 84% with a sensitivity of 83% and a specificity of 85% (Table 1). The LOO produced a concordance of 73% with a sensitivity of 72% and a specificity of 73% (Table 1) and the LMO for the Female Rat model produced a concordance of 70%.

3.3 External Validations

To simulate an external validation technique, 10% of the original learning set for each model was removed (Test Set) and the Rule Optimized model was identified using the remaining 90% of the compounds (Training Set). The most concordant model was used to predict the compounds in the Test set. Since the activities of all the compounds in the Test Set were known, we compared our predicted activities with the experimental activities and assessed whether our cat-SAR model was making an accurate prediction. The EXT validation process was repeated 10 times and overall average concordance, sensitivity, and specificity values were calculated.

The concordance values for the Complete Rat, Transgender Rat, Male Rat, and Female Rat models were 67%, 72%, 65%, and 69%, respectively (Table 2). These values were all similar to their respective LMO validation results (Table 1). The EXT analyses further validated the cat-SAR expert system. All of the results were significantly higher than randomness (50%), indicating that even with a large number of compounds removed from the learning set, enough information remained for the cat-SAR expert system to differentiate Rat, Male Rat, and Female Rat carcinogens from non-carcinogens.

3.4 Analysis of the Rat Models and the Role of Mutagenicity in Predicting Cancer

The values reported for each Rat model were close estimates to the experimental reproducibility of the rat cancer bioassays (85%) [43], and significantly higher than random chance predictions (50%). The results from the Complete Rat model were similar to those

reported by a previous study in which MultiCASE was used for SAR analysis of the CPDB [26]. However, after modification of the Rat dataset, cat-SAR significantly increased the concordance values from 69% to 74% ($p < 0.05$) (Table 1). Although several factors may be involved with this significant increase, one possibility is that it is linked to the accepted and general mechanism that carcinogenesis is caused through mutagenic effects to DNA. DNA-damaging compounds typically become electrophilic and contain biophores described by John Ashby as “structural alerts” for mutagenicity or carcinogenicity [18, 20]. Mutagenic carcinogens, therefore, share common structural characteristics and can be differentiated from non-mutagenic compounds through method such as SAR modeling. However, non-electrophilic/non-DNA binding (non-mutagenic) compounds cause tumorigenesis as well, possibly through receptor interactions and/or modifying critical cellular pathways that ultimately alter cellular functions.

Compounds that cause cancer in one gender and not the other may be doing so through a non-mutagenic mechanism since certain cellular targets may be gender specific. Analysis of the 97 carcinogens that were removed to create the Transgender Rat model demonstrated that only 55% of the compounds were correctly predicted by the Complete Rat model. Of these compounds, 78 of them had mutagenicity data, where 43 (55%) were non-mutagens and 35 (45%) were mutagens (data not shown). Furthermore, these compounds were also incorrectly predicted in the Male or Female Rat models as well, producing concordance values of 51% and 37%, respectively (data not shown). Furthermore, analysis of the 353 carcinogens in the Complete Rat model that had mutagenicity data demonstrated that 219 (62%) were mutagens and 134 (38%) were non-mutagens. After removal of the 78 carcinogens from the Transgender model, there were 275 carcinogens that had mutagenicity data and 184 (67%) were mutagens and 91 (33%) were non-mutagens (data not shown).

Investigation of the Complete Rat model demonstrated that of the 924 compounds, 612 compounds were tested in the Ames assay for *Salmonella* mutagenicity (295 were mutagenic and 317 were non-mutagenic) [27–33]. A summary of the distribution of compounds and their carcinogenic and mutagenic data is shown in Table 3. Further analysis of the cat-SAR predictions made on these compounds in all four Rat models showed how the cat-SAR expert system correctly predicted mutagenic carcinogens with a higher concordance than non-mutagenic carcinogens (Table 4). All four models correctly predicted the mutagenic carcinogens between 69 and 76% while only correctly predicting non-mutagenic carcinogens between 46 and 53% (Table 4). The multiple mechanisms of action through which non-mutagenic compounds have been proposed to produce tumorigenesis [45, 46] increase the variety of compound structures associated with non-mutagenic carcinogenicity, decreasing the common fragmental characteristics related to non-mutagenic effects.

3.5 Tumor Site Models

Another characteristic of non-mutagenic carcinogens is they typically cause cancer in only one tissue site. As opposed to mutagens which target DNA, non-mutagens may target cellular entities that may only be expressed in a single tissue site. Analysis of the 97 compounds removed to create the Transgender Rat model showed that there were 77

compounds (79%) that caused cancer in rats at one site, while 15 compounds (15%) caused cancer at two sites, 4 compounds (4%) caused cancer at 3 sites, and 1 compound (1%) caused cancer at 4+ sites. These results suggested that there may be a relationship between the number of sites that produce a tumor, mutagenicity and the concordance values of the cat-SAR rat cancer models. Therefore, several tumor site models were created by generating learning sets that consisted of carcinogens that produced cancer in either 1, 2, or 4+ sites compared to non-carcinogens.

Analysis of the LOO validations for the tumor site models demonstrated that as the number of sites increased the concordance between experimental and predicted values increased. The concordance values for the 1-site, 2-sites, and 4+-sites tumor models were 66%, 71%, and 80%, respectively (Table 4). The 4+-site model produced a higher concordance than any of the Rat, Male Rat, and Female Rat models. Furthermore, the number of mutagenic carcinogens and non-mutagenic carcinogens were analyzed for each of the tumor site models. The proportion of mutagenic carcinogens and non-mutagenic carcinogens increased as the number of tumor sites increased with the 1-site model having a proportion of 93 to 80 (54% mutagenic), the 2-sites model having a proportion of 60 to 28 (68% mutagenic) and the 4+-sites model having a proportion of 53 to 13 (80% mutagenic) (Table 4). These findings suggest that the mutagenic carcinogens are increasing the robustness of the rat cancer fragment model, and thus further suggest that non-mutagenic compounds have a much lower probability of being accurately predicted by SAR cancer models.

3.6 Demonstration Chemicals/Fragments Analysis

The cat-SAR expert system differentiates carcinogens from non-carcinogens through identification of structural attributes associated with carcinogenesis. Although there are various mechanisms through which compounds produce cancer, the Rat, Male Rat, and Female Rat models may be more suitable for correctly predicting mutagens as opposed to non-mutagens. As previously described through John Ashby's SAs, there are certain biophores that are associated with both mutagenicity and carcinogenicity. As a demonstration of how the cat-SAR expert system accurately detected these biophores, three compounds and the fragments that correctly predicted them as carcinogens were identified after analysis of the Transgender Rat model LOO validation.

The first compound analyzed was 3,3'-dimethylbenzidine (CAS# 612-82-8) (Figure 1), which is a *Salmonella* mutagen with a rat TD₅₀ of 0.629 mg/kg/day. The carcinogen is listed in the CPDB to target the rat Zymbal gland, large and small intestine, liver, lung, oral cavity, peritoneal cavity, preputial gland, skin, clitoral gland, and mammary gland [27–33]. There were 80 fragments used to correctly predict 3,3'-Dimethylbenzidine as a carcinogen, which were divided into two groups (Figure 1). Group 1 contained 7 fragments and represented the segment of the compound in which the amino group is attached to the benzene ring. This arylamine group has been previously described by Ashby as a SA for carcinogenicity [18, 20]. Furthermore, the mechanism by which the aminoaryl group is transformed into a DNA-reactive species and thus forms DNA adducts has been outlined by Verna et al [47]. The compound used in that study, 2-acetylaminofluorene (CAS# 53-96-3), is found in the Transgender Rat model learning set and was predicted as a carcinogen by the same

important fragments as 3,3'-dimethylbenzidine. Group 2 contained 73 fragments and represented the connection between the two aromatic rings. These fragments were found in 81% active compounds thus adding to the carcinogenic prediction of 3,3'-dimethylbenzidine.

The second compound analyzed, N-nitrosopropoxur (CAS# 38777-13-8) (Figure 2), is a derivative of the N-methylcarbamate insecticide, propoxur. This carcinogen is classified in the CPDB as a *Salmonella* mutagen with a rat TD₅₀ of 0.364 mg/kg/day, causing cancer in the male rat stomach [27–33]. Although propoxur is not a known mutagen, the addition of the nitroso functional group has been shown to cause DNA damage in hamster lung fibroblasts, rat tracheal epithelial cells and human gastric cells [48, 49]. In fact, the N-nitroso-N-methyl functional group is described by Ashby as one of his SA's for mutagenic carcinogens [18, 20]. There were 36 fragments used to correctly predict N-nitrosopropoxur as a rat carcinogen, which were divided into four representative groups. Of the four representative fragments, three contained a portion of the N-nitroso-N-methyl functional group (Groups 1–3 Figure 2), and all were found in at least 88% active compounds. The fourth and final group lacked the Ashby SA and was only found in 1 active compound but 5 inactive compounds (Group 4, Figure 2).

The third compound analyzed, β -propiolactone (CAS# 57-57-8) (Figure 3) is a *Salmonella* mutagen with a rat TD₅₀ of 1.46 mg/kg/day in rats. This carcinogen is listed in the CPDB as targeting the female rat stomach [27–33]. β -Propiolactone, which contains a cyclic ester, was shown *in vitro* to produce adducts with nucleic acids [50]. The entire structure of β -Propiolactone is considered one of Ashby's SA's for mutagenic carcinogens [18, 20]. There were two fragments that correctly predicted β -Propiolactone as a rat carcinogen, which were divided into two representative groups (Groups 1 and 2, Figure 3). Both groups of fragments were found in 100% active compounds and contain the cyclic ester functional group (Figure 3).

4. Conclusion

All three Rat, Male Rat and Female Rat models produced similar concordance values with the Transgender Rat model producing the highest value of 74% (Table 1). The goal of separating compounds that affected one gender and not the other was to identify specific fragments that were associated with gender-specific carcinogenesis as well as to study how well the cat-SAR carcinogenesis models perform when mutagenicity is affecting the model. Gender specific effects may be most likely produced through non-mutagenic mechanisms rather than through directly damaging DNA (or they may simply be from a misclassification in one gender). This study demonstrated that the cat-SAR Rat models correctly predicted the mutagenic compounds with a higher concordance than the non-mutagenic compounds. Furthermore, when analyzing the effect of producing cancer at multiple sites on the predictivity of cat-SAR models, it was observed that as the number of tumor sites increased so did the model's concordance values. In fact, the 4+-sites model produced a concordance value of 80% which was the highest of all of the rat cancer models. Moreover, there is a positive correlation between the number of tumor sites and the number of mutagenic carcinogens, which further strengthens the notion that fragment-based SAR models lack the ability to accurately predict non-mutagenic carcinogens. Analysis of fragments from the

Transgender Rat model LOO validation demonstrated that the common active biophores were similar to the SA's identified in past studies by John Ashby, reinforcing the idea that biophores associated with cancer are strongly associated to mutagenicity.

In conclusion, future SAR models for the predictions of carcinogens may need to be a two-step process. There will have to be an understanding that predicting non-mutagenic compounds with current SAR fragment models may lead to an accuracy of around only 50% as demonstrated in our cat-SAR modeling results. Therefore, a result from an *in vitro* Ames assay for *Salmonella* mutagenicity or even a prediction of mutagenicity using an SAR model based on *Salmonella* mutagenicity data may need to be the first step before making a prediction for carcinogenicity. Based on the *Salmonella* results, the second step would be split into two directions. As we have previously reported, the cat-SAR model consisting of non-mutagenic carcinogens and non-carcinogens produced a concordance of 71%. Therefore, receptor-based cat-SAR models should be used to predict non-mutagens as potential carcinogens. Mutagenic compounds, however, should be predicted using fragment-based cat-SAR models such as the Transgender Rat model or the 4+-site model. These models decrease the number of non-mutagens in the model and thus increase the likelihood of a more accurate cancer prediction.

Acknowledgments

This research was supported by the National Institutes of Health (P20 RR018733) and the James Graham Brown Cancer Center, University of Louisville.

References

1. U. S. EPA. [last accessed 11/11/13] TSCA Inventory Reset. 2010. Available at <http://www.epa.gov/champ/pubs/hpv/tsca.html>
2. Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E. The toxicity data landscape for environmental chemicals. *Environ Health Perspect.* 2009; 117:685–695. [PubMed: 19479008]
3. Huff J. Long-term chemical carcinogenesis bioassays predict human cancer hazards. Issues, controversies, and uncertainties. *Ann NY Acad Sci.* 1999; 895:56–79. [PubMed: 10676409]
4. P. European and Council. Regulation (EC) No 1907/2006 of the European Parliament and of the Council. *Official Journal of the European Union* L396. 2007
5. Williams ES, Panko J, Paustenbach DJ. The European Union's REACH regulation: a review of its history and requirements. *Crit Rev Toxicol.* 2009; 39:553–575. [PubMed: 19650717]
6. Koop BF. Human and rodent DNA sequence comparisons: a mosaic model of genomic evolution. *Trends Genet.* 1995; 11:367–371. [PubMed: 7482789]
7. Huff J. Animal and human carcinogens. *Environ Health Perspect.* 1999; 107:A341–A342. [PubMed: 10405250]
8. Ames BN, Gold LS. Chemical carcinogenesis: too many rodent carcinogens. *Proc Natl Acad Sci USA.* 1990; 87:7772–7776. [PubMed: 2217209]
9. Ashby J. International Commission for Protection Against Environmental Mutagens and Carcinogens. Two million rodent carcinogens? The role of SAR and QSAR in their detection. *Mutat Res.* 1994; 305:3–12. [PubMed: 7508545]
10. Krewski D, Andersen ME, Mantus E, Zeise L. Toxicity testing in the 21st century: implications for human health risk assessment. *Risk Anal.* 2009; 29:474–479. [PubMed: 19144067]
11. Benigni R, Bossa C. Predictivity of QSAR. *J Chem Inf Model.* 2008; 48:971–980. [PubMed: 18426198]

12. Benigni R, Bossa C. Predictivity and reliability of QSAR models: the case of mutagens and carcinogens. *Toxicol Mech Methods*. 2008; 18:137–147. [PubMed: 20020910]
13. 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]
14. Rosenkranz HS, Klopman G. Mutagens, carcinogens, and computers. *Prog Clin Biol Res*. 1986; 209A:71–104. [PubMed: 3529111]
15. Rosenkranz HS, Klopman G. Structural basis of carcinogenicity in rodents of genotoxicants and non-genotoxicants. *Mutat Res*. 1990; 228:105–124. [PubMed: 2300064]
16. Blake BW, Enslein K, Gombar VK, Borgstedt HH. Salmonella mutagenicity and rodent carcinogenicity: quantitative structure-activity relationships. *Mutat Res*. 1990; 241:261–271. [PubMed: 2195334]
17. Marchant CA. Prediction of rodent carcinogenicity using the DEREK system for 30 chemicals currently being tested by the National Toxicology Program. The DEREK Collaborative Group. *Environ Health Perspect*. 1996; 104(Suppl 5):1065–1073. [PubMed: 8933056]
18. Ashby J. Fundamental structural alerts to potential carcinogenicity or noncarcinogenicity. *Environ Mutagen*. 1985; 7:919–921. [PubMed: 4065064]
19. Ashby J, Paton D. The influence of chemical structure on the extent and sites of carcinogenesis for 522 rodent carcinogens and 55 different human carcinogen exposures. *Mutat Res*. 1993; 286:3–74. [PubMed: 7678908]
20. Ashby J, Tennant RW. Chemical structure, Salmonella mutagenicity and extent of carcinogenicity as indicators of genotoxic carcinogenesis among 222 chemicals tested in rodents by the U.S. NCI/NTP. *Mutat Res*. 1988; 204:17–115. [PubMed: 3277047]
21. Zhu H, Rusyn I, Richard A, Tropsha A. Use of cell viability assay data improves the prediction accuracy of conventional quantitative structure-activity relationship models of animal carcinogenicity. *Environmental health perspectives*. 2008; 116:506–513. [PubMed: 18414635]
22. Benigni R, Zito R. The second National Toxicology Program comparative exercise on the prediction of rodent carcinogenicity: definitive results. *Mutat Res*. 2004; 566:49–63. [PubMed: 14706511]
23. Valerio LG Jr. In silico toxicology for the pharmaceutical sciences. *Toxicol Appl Pharmacol*. 2009; 241:356–370. [PubMed: 19716836]
24. Zeeman M, Auer CM, Clements RG, Nabholz JV, Boethling RS. U.S. EPA regulatory perspectives on the use of QSAR for new and existing chemical evaluations. *SAR QSAR Environ Res*. 1995; 3:179–201. [PubMed: 8564854]
25. U. S. EPA. [last accessed 11/11/13] OncoLogic™ - A Computer System to Evaluate the Carcinogenic Potential of Chemicals. 2013. Available at <http://www.epa.gov/oppt/sf/pubs/oncologic.htm>
26. Cunningham AR, Klopman G, Rosenkranz HS. Identification of structural features and associated mechanisms of action for carcinogens in rats. *Mutat Res*. 1998; 405:9–27. [PubMed: 9729240]
27. Gold LS, de VM, Backman GM, Magaw R, Lopipero P, Smith M, Blumenthal M, Levinson R, Bernstein L, Ames BN. Chronological supplement to the Carcinogenic Potency Database: standardized results of animal bioassays published through December 1982. *Environ Health Perspect*. 1986; 67:161–200. [PubMed: 3530736]
28. Gold LS, Manley NB, Slone TH, Garfinkel GB, Rohrbach L, Ames BN. The fifth plot of the Carcinogenic Potency Database: results of animal bioassays published in the general literature through 1988 and by the National Toxicology Program through 1989. *Environ Health Perspect*. 1993; 100:65–168. [PubMed: 8354183]
29. Gold LS, Manley NB, Slone TH, Rohrbach L. Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature in 1993 to 1994 and by the National Toxicology Program in 1995 to 1996. *Environ Health Perspect*. 1999; 107(Suppl 4):527–600. [PubMed: 10421768]
30. Gold LS, Manley NB, Slone TH, Rohrbach L, Garfinkel GB. Supplement to the Carcinogenic Potency Database (CPDB): results of animal bioassays published in the general literature through 1997 and by the National Toxicology Program in 1997–1998. *Toxicol Sci*. 2005; 85:747–808. [PubMed: 15800034]

31. Gold LS, Sawyer CB, Magaw R, Backman GM, de VM, Levinson R, Hooper NK, Havender WR, Bernstein L, Peto R. A carcinogenic potency database of the standardized results of animal bioassays. *Environ Health Perspect.* 1984; 58:9–319. [PubMed: 6525996]
32. Gold LS, Slone TH, Backman GM, Eisenberg S, Da CM, Wong M, Manley NB, Rohrbach L, Ames BN. Third chronological supplement to the carcinogenic potency database: standardized results of animal bioassays published through December 1986 and by the National Toxicology Program through June 1987. *Environ Health Perspect.* 1990; 84:215–286. [PubMed: 2351123]
33. Gold LS, Slone TH, Backman GM, Magaw R, Da CM, Lopipero P, Blumenthal M, Ames BN. Second chronological supplement to the Carcinogenic Potency Database: standardized results of animal bioassays published through December 1984 and by the National Toxicology Program through May 1986. *Environ Health Perspect.* 1987; 74:237–329. [PubMed: 3691431]
34. Cunningham AR, Carrasquer CA, Mattison DR. A categorical structure-activity relationship analysis of the developmental toxicity of antithyroid drugs. *Int J Pediatr Endocrinol.* 2009; 2009:936154. [PubMed: 20111734]
35. Cunningham AR, Cunningham SL, Consoer DM, Moss ST, Karol MH. Development of an information-intensive structure-activity relationship model and its application to human respiratory chemical sensitizers. *SAR QSAR Environ Res.* 2005; 16:273–285. [PubMed: 15804814]
36. 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]
37. DSSTox, Distributed Structure-Searchable Toxicity (DSSTox) Database Network. [last accessed 11/11/13] CPDBAS: Carcinogenic Potency Database Summary Tables - All Species Database File, version 5D. 2008. Available at http://www.epa.gov/ncct/dsstox/sdf_cpdbas.html
38. Zeiger E. Historical perspective on the development of the genetic toxicity test battery in the United States. *Environmental and molecular mutagenesis.* 2010; 51:781–791. [PubMed: 20740645]
39. Combes RD. The use of structure-activity relationships and markers of cell toxicity to detect non-genotoxic carcinogens. *Toxicology in vitro : an international journal published in association with BIBRA.* 2000; 14:387–399. [PubMed: 10906445]
40. 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. *Expert Opinion in Drug Metabolism and Toxicology.* 2010; 6:1–11.
41. Woo, YT. Mechanisms of action of chemical carcinogens, and their role in Structure-Activity Relationships (SAR) analysis and risk assessment. In: Benigni, R., editor. *Quantitative Structure-Activity Relationship (QSAR) Models of Mutagens and Carcinogens.* CRC Press; Boca Raton: 2013. p. 41-80.
42. Cunningham AR, Carrasquer CA, Qamar S, Maguire JM, Cunningham SL, Trent JO. Global structure-activity relationship model for nonmutagenic carcinogens using virtual ligand-protein interactions as model descriptors. *Carcinogenesis.* 2012; 33:1940–1945. [PubMed: 22678118]
43. 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.
44. 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]
45. Shaw IC, Jones HB. Mechanisms of non-genotoxic carcinogenesis. *Trends Pharmacol Sci.* 1994; 15:89–93. [PubMed: 8184492]
46. Yamasaki H. Non-genotoxic mechanisms of carcinogenesis: studies of cell transformation and gap junctional intercellular communication. *Toxicol Lett.* 1995; 77:55–61. [PubMed: 7618169]
47. Verna L, Whysner J, Williams GM. 2-Acetylaminofluorene mechanistic data and risk assessment: DNA reactivity, enhanced cell proliferation and tumor initiation. *Pharmacol Ther.* 1996; 71:83–105. [PubMed: 8910950]
48. Kuo HH, Shyu SS, Wang TC. Genotoxicity of low dose N-nitroso propoxur to human gastric cells. *Food Chem Toxicol.* 2008; 46:1619–1626. [PubMed: 18329776]

49. Wang TC, Chiou JM, Chang YL, Hu MC. Genotoxicity of propoxur and its N-nitroso derivative in mammalian cells. *Carcinogenesis*. 1998; 19:623–629. [PubMed: 9600347]
50. Hemminki K. Reactions of beta-propiolactone, beta-butyrolactone and gamma-butyrolactone with nucleic acids. *Chemico-biological interactions*. 1981; 34:323–331. [PubMed: 6161710]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

3,3'-dimethylbenzidine
CAS# 612-82-8

Group 1
12/13
0.92

Group 2
30/37
0.81

Figure 1.
Illustration of the fragments predicting 3,3'-dimethylbenzidine as a rat carcinogen.

N-nitrosopropoxur
CAS# 38777-13-8

Figure 2.
Illustration of the fragments predicting N-nitrosopropoxur as a rat carcinogen.

β -Propiolactone
CAS# 57-57-8

Group 1
2/0
1.00

Group 2
2/0
1.00

Figure 3.
Illustration of the fragments predicting β -Propiolactone as a rat carcinogen.

Table 1

Fragment summary, self-fit, and cross validation results for all rat carcinogen – non-carcinogen cat-SAR models.

Model	Fragments			Leave None Out			Leave One Out			Leave Many Out		
	Total	Model	Active	Inactive	Sensitivity	Specificity	Concordance	Sensitivity	Specificity	Concordance	Sensitivity	Specificity
Complete Rat	43557	36330	19456	16874	86% (418/486)	86% (297/346)	86% (715/832)	70% (333/478)	68% (237/349)	69% (570/827)	69% (31.25/45.41)	67% (22.16/33.32)
Transgender Rat	40348	32990	13791	19199	90 (354/394)	89 (320/358)	90 (674/52)	73 (287/392)	74 (258/350)	74 (545/742)	73 (26.82/36.84)	73 (24.44/33.27)
Male Rat	42040	14550	6202	8348	83 (327/396)	76 (290/381)	79 (617/777)	66 (259/390)	67 (257/382)	67 (516/772)	65 (24.08/36.99)	68 (24.82/36.47)
Female Rat	42009	11180	3888	7292	83 (271/327)	85 (355/416)	84 (626/743)	72 (245/341)	73 (302/413)	73 (547/754)	71 (22.80/32.49)	69 (27.04/39.03)

SAR QSAR Environ Res. Author manuscript; available in PMC 2016 April 13.

Table 2

External validation summary of results for the Rat, Male Rat and Female Rat cat-SAR models. The values are calculated by averaging the sensitivity, specificity, and concordance of 10 random sets.

Model	Sensitivity	Specificity	Concordance
Complete Rat	69±8% (31/45)	66±6% (25/38)	67±4% (56/83)
Transgender Rat	71±7 (29/41)	74±7 (26/35)	72±4 (55/76)
Male Rat	65±10 (26/40)	65±9 (26/40)	65±6 (52/80)
Female Rat	68±8 (25/37)	70±6 (28/40)	69±4 (53/77)

Table 3

Number of carcinogens and non-carcinogens with *Salmonella* mutagenesis data, and the number of correct predictions for these compounds identified from analysis of the Complete Rat Model LOO validation.

	Carcinogen		Non-Carcinogen		Total Compounds
	Number of Compounds	Correct Predictions	Number of Compounds	Correct Predictions	
Mutagen	219	75% (155/208)	76	57% (39/68)	295
Non-Mutagen	134	46 (51/111)	183	70 (116/165)	317
Total	353		259		612

Table 4

Effect of carcinogenicity and mutagenicity on accurately predicting rat carcinogens.

Model	Carcinogens Mutagens	Non-carcinogens Non-mutagens	Overall Concordance
Complete Rat	75% (150/199)	47% (49/105)	65% (199/304)
Gender Concordant Rat	76 (126/165)	48 (35/73)	68 (161/238)
Male Rat	69 (113/163)	46 (46/100)	60 (159/263)
Female Rat	76 (117/154)	53 (41/78)	68 (158/232)

Table 5

Effect of increasing the number of tumor sites on accurately predicting rat carcinogens and the number of Mutagenic Carcinogens and Non-Mutagenic Carcinogens.

Model	Sensitivity	Leave One Out Specificity	Concordance	Mutagenic Carcinogens	Non-Mutagenic Carcinogens
Single Site	66% (167/253)	67% (165/248)	66% (332/501)	93	80
Two Site	71 (65/92)	71 (63/89)	71 (128/181)	60	28
Four or More Sites	79 (62/78)	80 (64/80)	80 (126/158)	53	13