

## **Predictive Toxicology Challenge (PTC) 2000-2001:** A Toxicologist's View and Evaluation

Yin-tak Woo  
Risk Assessment Division (7403M)  
Office of Prevention, Pesticides and Toxic Substances  
U.S. Environmental Protection Agency  
1200 Pennsylvania Avenue, NW  
Washington, DC 20460  
[woo.yintak@epa.gov]

### **"Ideal" Predictive System(s)**

Predictive toxicology is a multi-disciplinary science that requires the close collaboration among toxicologists, chemists, biologists, statisticians and AI/machine learning researchers. Since the final practical end-users of most predictive systems are mainly toxicologists (involved in hazard identification and risk assessment) and chemists (involved in design and development of safer and effective chemicals), it is imperative and useful for scientists in the other disciplines to be aware of the specific needs of chemists and toxicologists in order to develop the most effective predictive system(s). From my decades of experience in development and evaluation of predictive systems as a toxicologist, "ideal" predictive system(s) should: (a) be flexible and capable of being optimized for specific toxicological endpoint and specific types of chemical compounds when necessary, (b) incorporate all available, relevant information including chemical structural, biological and exposure scenario, (c) derive from statistical, intuitive or relational association with mechanistic backing, (d) be trained using learning/knowledge databases with adequate representation of chemicals structurally and/or functionally related to the test chemicals, (e) contain provisions to estimate its own predictive boundary, limitation and uncertainty, (f) be able to generate hypothesis for testing, (g) provide rationale for prediction and/or analogs/surrogates for comparison, (h) be interpretable and conducive to molecular design applications, and (i) provide some indication of relative potency or trans-species potential.

### **Difficulties of Predicting Carcinogenic Activity**

To assess the possible outcome of PTC 2000-2001 project, it is important to be aware of the difficulties of predicting carcinogenic activity. Owing to the high cost of long-term cancer bioassays and the availability of structure-activity relationships (SAR) data for some classes of chemicals, prediction of carcinogenic activity of chemicals has been the focus of numerous studies. Whereas significant progresses have been made in some well defined structural classes of chemicals, the prediction of carcinogenic activity of a random or heterogeneous set of chemicals remains one of the most difficult and

challenging problems. Some of major difficulties in this endeavor include: (a) mechanistic complexity involving multi-stage, multi-factorial etiology, (b) insufficient training knowledge/database, (c) limitation of reproducibility of cancer bioassay data, (d) difficulty of feedback/hypothesis testing/generation in model development.

### **Inherent Limitations of the PTC 2000-2001 Project**

The PTC 2000-2001 project basically used the cancer bioassay data of close to 500 National Toxicology Program (NTP) chemicals as the learning/training database to predict retrospectively the results of 185 Food and Drug Administration (FDA) pharmaceutical chemicals. Beyond the difficulties described above, there are also some inherent limitations of the PTC 2000-2001 project that may limit the predictive capability. These include: (a) differential coverage of chemicals in the learning/training set (a heterogeneous database with industrial chemicals, pharmaceuticals, dyes and polymeric substances of diverse chemical structures) *versus* test chemical set of only pharmaceutical chemicals, (b) consideration of chemical structures only, (c) consideration of the qualitative aspects of the bioassay data only.

### **Evaluation of the Toxicological Significance of PTC Models**

Owing to time constraint, the following evaluation is mainly based on review of the extended abstracts of the PTC models. Assignment of toxicological significance to useful predictive rules/descriptors/attributes may be subjective and therefore subject to modification if detailed information on the chemical compounds that contribute to the derivation of the rule/descriptor is available.

1. Overall view: Overall, my impression is that this is an interesting study with some innovative approaches. A number of descriptors or approaches may prove to be useful in predicting carcinogenicity and/or providing insights for further studies. Although the predictive outcome of the test set does not seem to be too impressive, it is to be expected from heterogeneous database on an endpoint with such complex mechanism(s). Based on review of the abstracts, there seems to be a lack of discernible knowledge discovery although this may change after full details become available. Some of the findings seem to be difficult to interpret from a toxicological viewpoint. There seems to be insufficient interaction between the model developers and toxicologists. With some exception, the handling of boundary, limitation and uncertainty is also less than satisfactory
2. Optimal Model by Blinova et al. (VINITD): In this paper, the authors used *fragmentary code of substructure superposition* (FCSS) as the descriptor language to describe chemical compounds. A chemical compound is described as a set of substructures that are centers of localization of  $\pi$ -electrons (active or *descriptor centers*) separated by chains of carbon pairs. Toxicology analysis was conducted using the simple JSM-method (which involves formal concept analysis) to identify structural attributes associated with carcinogenicity. Apart from detailed description of the FCSS methodology, the abstract itself is not very informative.

However, independent evaluation by Pfahringer (see item 4 below) listed the method as being the best (out of seven PTC features) in contributing to total and unique counts of "carcinogenic" predictions from test set. Other than capturing the informational content of chemical compounds, the toxicological significance of the FCSS method is not clear. The importance of having conjugation between carbon chains connecting descriptor centers is consistent with toxicological mechanistic consideration that conjugation may contribute to resonance stabilization of metabolically activated reactive intermediates to enhance opportunity for interaction with target macromolecules.

3. Optimal Model by Blockeel et al. (Leuven): This paper described the development of first order decision tree for the prediction of carcinogenic activity based on the presence/absence of Br, O, S, and Cl. Four sets of nested if-then-else rules, one for each animal and gender set (male mouse, female mouse, male rat, female rat) were developed. Based on training from NTP database, the Br rule is considered so strong that the authors always predict POSITIVE when a bromine is present. The high potential of brominated compounds to be carcinogenic is consistent with toxicological knowledge that bromine is a good leaving group and therefore could generate reactive intermediates from brominated compounds. It is not clear whether the authors intend to set predictive domain or boundary for the rule. It would not be reasonable to expect all brominated compounds to be carcinogenic. There are no details on "thresholds" for O, S and Cl.
4. Optimal Model by Pfahringer (Waikato): This paper provided an objective evaluation of the discriminating ability of the optimized classifiers from seven different sets of PTC features and used a voting approach to optimally combine these sets into an overall prediction scheme. Using male rat carcinogenicity as the predicted endpoint, the author found that model M1 (which predicts "carcinogenic" if at least one of the seven classifiers says so) performed the best. Nevertheless, the predictive accuracy was judged to be still less than satisfactory. The analysis of the classifier contributions to total and unique counts of "carcinogenic" predictions for each feature set is interesting and may be of toxicological significance in finding optimal combination of complementary methods and in providing mechanistic insights.
5. Optimal Model by Gonzalez et al. (Uta): This model used a set of 24 substructures to predict the carcinogenic potential of chemicals in the male rat. The substructures were applied in order. As soon as one substructure was found in the compound, a positive prediction would be made with no further analysis. A compound would be predicted to be negative if it did not contain any of the 24 substructures. There was no information on the predictive confidence of each of the substructures and on the basis for the hierarchy. The toxicological significance of some of the substructures (e.g., #1 for dibromoalkanes, #7 for aromatic amines with ring methyl substitution) was obvious for established structural classes of carcinogens; however, there was no provision for positional

effects (e.g., vicinally *versus* geminally substituted bromine; *ortho*-methyl *versus* other ring position) and boundaries. Others appeared to be quite vague (e.g., #6, fused or non-fused ring? #11, methyl esters?) and would require examination of compounds contributing to the generation of the discriminating substructure. Since the 24 substructures could not possibly cover the whole universe of chemical carcinogens, the confidence that could be placed on negative predictions may be debatable. It is interesting to note that only 7 of the 24 substructures could be used on the test chemical set implying differential coverage of compounds in the training *versus* test set. The dominance of substructure #7 is consistent with the expected structural dominance of aromatic amines in pharmaceuticals.

6. Optimal Model by Okada (Kwansei): This paper used the cascade model as a rule induction methodology for finding characteristic substructures and properties that are useful for predicting chemical carcinogenicity. The method offers an interesting and effective means to set main condition and precondition and allow combination of discriminating characteristics into rules. The strength of each rule can be analyzed by between-group sum of squares analysis. By analyzing change in probability, some rules can also be used to provide valuable insight for further studies. Among the "strong" rules identified, the absence of O (HBA = 0) plus high flexibility (FLEX > 0.5) was the strongest. The toxicological significance is unclear. The absence of O may be considered favorable in terms of no having hydrophilic groups such as -OH but would not work for compounds like epoxides. High molecular flexibility is favorable for cross-linking agents but will not necessarily work for polynuclear compounds. As in previous method, the lack of applicability of strong rules derived from the training set to the test set indicated differential coverage of chemicals in the two sets..

### **Approaches to *Post-mortem* Analysis**

The most important lesson one can learn from predictive exercises such as PTC 2000-2001 is through analyses of what worked well and what went wrong. The following are suggestions of possible approaches. (a) analysis of false positives and false negatives, (b) study of the role of mechanism in affecting prediction, (c) comparison of class representation of chemicals in the learning set and test set, (d) possible class partitioning of learning and test sets, (e) expansion of allowable descriptors.

**Disclaimer:** The scientific opinions expressed in this abstract are solely those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency. Mention of tradenames does not necessarily signify endorsement by the U.S. Government.