Development of Knowledge Rules for Cancer Expert System for Prediction of Carcinogenic Potential of Chemicals: USEPA Approach

Yin-tak Woo, David Y. Lai, Joseph C. Arcos (*ret.)*, and Mary F. Argus (*ret.)*

Office of Pollution Preventions and Toxics (MC 7403), U.S. Environmental Protection Agency, 401 M Street SW, Washington, DC 20460

Since the passage of the Toxic Substances Control Act (TSCA), which requires the U.S.Environmental Protection Agency (EPA) to assess potential toxic effects of new and existing chemicals, structure-activity relationships (SAR) analysis has been extensively and effectively used in the identification of potential health hazards of new (Premanufacturing Notification) industrial chemicals (Arcos, 1983; Auer and Gould, 1987). To evaluate the potential health hazards of the large number of existing chemicals for which adequate test data are not available, SAR analysis has been given an increasing role as a basis for recommending additional testing, for designing strategic research plans in TSCA programs, and for setting priorities of testing environmental pollutants such as disinfection byproducts in drinking water (Woo et al., 1999). To carry out pollution prevention initiatives, SAR has also been used to search for safer chemical substitutes and in molecular design of new chemicals (DeVito and Garrett, 1996). Most recently, SAR was used for the first time to provide support for the regulatory decision to list 2,4,6-tribromophenol as a hazardous waste (EPA, 1998).

Most of the SAR analysis in TSCA programs has been carried out by a team of experts (the Structure Activity Team) with extensive knowledge of chemistry, toxicology and risk assessment. To capture the expertise of the Structure Activity Team, the Agency started a collaborative project with LogiChem, Inc. to develop a knowledge rule-based expert system, the OncoLogic system, to predict the carcinogenic potential of chemicals. In this paper, we present our approaches for the development of knowledge rules for the system.

The use of knowledge rules has several advantages: (i) they are easy to visualize and understand so that users may be able to better appreciate the reasoning and even apply to chemicals of their own interest , (ii) they are amenable to mechanistic interpretation so that the basis for concern can be strengthened or justified, (iii) they are amenable to hypothesis testing so that uncertainties or knowledge gaps can be filled by designing strategic research, and (iv) they can be utilized for molecular design of safer chemicals.

There are many possible approaches to the discovery and development of knowledge rules; these include: (i) automated data mining, (ii) correlative/statistical, (iii) mechanistic, (iv) bioavailability-based, (v) metabolism-based, (vi) exposure-based, and (vii) policy-based considerations. Ideally, rules discovered from automatic data mining or correlative analysis should have mechanistic backing for effective justification.

The main approach used by EPA's Structure Activity Team is mechanism-based with consideration of bioavailability, metabolism, exposure and policy. Computer-assisted substructure searches are also used to link all relevant data together. The basic principles used in mechanism-based SAR analysis of chemical carcinogens have been discussed (Woo et al., 1995, Lai and Woo, 1999). Essentially, from the systematic review of over 12,000 chemical carcinogens of various structural classes covered in a series of monographs (Arcos and Argus, 1974a,b, Arcos et al., 1982, Woo et al., 1985,1988) and various other new sources of information, a number of common features are discernible. The profile of typical potent carcinogens includes: (i) ability to reach target sites (bioavailability and exposure consideration), (ii) activation, or staying active, near or at target site (exposure and metabolism consideration), (iii) reasonable lifetime to allow interaction with target macromolecules (resonance stabilization of reactive intermediate or metabolically persistent receptor ligand), (iv) selective, specific and persistent interaction with target macromolecules, and (v) ability to act on various stages--initiation, promotion and progression--of carcinogenesis (mechanistic consideration). These principles are basically applicable to all structural classes of compounds. To maximize predictive accuracy, however, structure class-specific knowledge rules have been developed to address the special features unique to the different classes.

For each structural class, the process for rule discovery and development involve: (i) gathering all available information, (ii) brainstorming to determine mechanism-based structural features which contribute to carcinogenicity and secondary factors which modify

activity, (iii) determining the need to subclassify into sub-classes to optimize predictive accuracy, and (iv) narrowing down to crucial factors, assigning importance to each factor and evaluating their suitability for system development.

To illustrate the process, we use aromatic amines as an example. Aromatic amines represent one of the most well established class of chemical carcinogens of occupational and environmental importance with numerous studies (Woo and Lai, 1999). The basic requirements for an aromatic amine to be carcinogenic are: the presence of an aromatic ring system (a single ring or more than one ring forming a conjugated system, fused or non-fused) and of the amine/amine-generating group (owing to the possible metabolic interconversion of the amino group with hydroxylamino and nitroso groups and the metabolic reduction of the nitro to nitroso, all these groups are often termed "amine-generating groups"). Figure 1 shows the aromatic ring systems in some well-known carcinogenic aromatic amines. The unattached bonds on these ring systems indicate the positions where attachment of amine or amine-generating group gives rise to the most potent carcinogenic compounds. These positions are almost invariably the terminal end of the longest conjugated chain in the aryl moiety. As shown in Figure 2, such positions are most effective in resonance stabilization of reactive nitrenium ions which could be generated after metabolic activation. Mechanistically, resonance stabilization apparently provides extended lifetime for reactive intermediates to travel from the site of metabolic activation to target macromolecules to initiate carcinogenesis.

Beyond the presence of an aromatic ring system and of amine/amine-generating group(s), SAR analysis has revealed several other structural features important for predicting the carcinogenicity of aromatic amines. These include: (i) number and nature of aromatic rings; (ii) nature of amine/amine-generating group(s); (iii) nature, number and position of other ring substituent(s); and (iv) size, shape and planarity of the molecules. Figure 3 shows a synoptic summary of the effect of ring substitutions on the carcinogenic activity of 4-aminobiphenyl and benzidine. Of particular interest are the observations that the reduction of carcinogenic activity by introduction of hydrophilic or bulky substituents or substitutents at the 2,2',6,6'-position(s) which may distort molecular planarity. From these data and other supportive information, a number of knowledge rules with regard to aromatic amine carcinogenicity have been developed; these include: (i) amino group occupying the terminal end of longest conjugated chain most favorable, (ii) disruption of resonance stabilization inhibitory, (iii) amino groups with alkyl substitution beyond methyl inhibitory especially if branched, (iv) ring substitution with bulky groups flanking amino group inhibitory, (v) ring substitution with highly hydrophilic groups inhibitory, (vi) ring substitution(s) that distort planarity inhibitory, and (vii) highly elongated ring system unfavorable.

These rules have been used in a prospective prediction of the outcome of NTP bioassays before they became available. As can be seen from the table in Figure 5, these rules have accurately predicted the negative carcinogen with a concern level of Low, and the clearly active carcinogens with concern level of High-Moderate. In between, we have predicted Marginal or Low-Moderate concerns for compounds which were shown to be weakly carcinogenic or have equivocal results. Thus, at least for this batch of chemicals, these rules have done a reasonably good job of predicting not only the carcinogenic activity but also their relative potency. An expanded prediction of another 30 NTP chemicals has been published (Woo et al., 1997).

Knowing the factors which contribute to or reduce carcinogenic activity, it should be possible to synthesize, by molecular designing, safer chemicals by avoiding the factors which are favorable and capitalizing on factors which are unfavorable for conferring carcinogenic activity. Figure 6 shows a number of approaches that organic chemists may utilize to design safer chemicals. The underlying rationales for such approaches are also included. It is hoped that by designing safer chemicals before they are introduced to commercial uses, the chemical industries can make significant contribution to pollution prevention.

It should be pointed out that purely structure-based SAR analysis does not always work effectively for all types of chemicals. For certain types of chemicals, incorporation of biological information may significantly enhance predictive accuracy. For example, it is very difficult to predict the carcinogenic potential of organophosphorus compounds because of their complex mechanisms, acute toxicity and rapid detoxification. By incorporating biological information, a better prediction is achievable (Woo et

al., 1996). For chemicals in general, an integrative approach of combining mechanistically complementary short-term predictive tests as a basis for assessing the carcinogenic potential of chemicals has been proposed (Woo et al., 1998).

The mechanism-based SAR analysis is a relatively resource intensive and time consuming process. It is hoped that the AI community may develop newer or improved technology for rule discovery and refinement, to improve the overall predictive accuracy and capability. In the process, however, they should be aware of the needs to incorporate mechanistic understanding and/or biological information whenever possible. Efforts should be spent to develop hybrid systems which complement each other to maximize their strengths and circumvent their individual limitations. Collaboration with toxicologists, and outreach to industrial and regulatory communities will ensure the development of truly useful toxicity predictive systems.

Acknowledgment: The authors wish to thank Dr. Marilyn Arnott for her contribution to the development of the OncoLogic system (website: www.logichem.com) and for invaluable input and feedback during refinement of our rules.

Disclaimer: The views expressed in this paper are solely those of the authors and do not necessarily reflect those of the Environmental Protection Agency. Mention of trade-names, commercial products or organization does not imply endorsement by the U.S. Government.

References

Arcos, J.C. 1983. Comparative Requirements for Premarketing/premanufacture Notifications in the EC Countries and the USA with Special Reference too Risk Assessment in the Framework of the U.S. Toxic Substances Control Act. *J. Amer. College Toxicol.* 2,131-145.

Arcos, J.C., and Argus, M.F. 1974a. *Chemical Induction of Cancer. Structural Bases and Biologic Mechanism. Vol. II A, Polynuclear aromatic hydrocarbons.* Academic Press, N.Y.

Arcos, J.C. and Argus, M.F. 1974b. *Chemical Induction of Cancer. Structural Bases and Biologic Mechanism. Vol. II B, Aromatic amines and related compounds.* Academic Press, N.Y..

Arcos, J.C. Woo, Y.-T., and Argus, M.F. (with the collaboration of DY Lai). 1982. *Chemical Induction of Cancer. Structural Bases and Biologic Mechanism. Vol. IIIA, Aliphatic carcinogens.* Academic Press, N.Y.

Auer, C.M., and Gould, D.H. 1987. Carcinogenicity Assessment and the role of Structure-Activity Relationship Analysis under TSCA Section 5. *Environ. Carcino. Revs.* C5, 29-71.

DeVito, S.C., and Garrett, R.L.eds. 1996. *Designing Safer Chemicals,* ACS Symposium Ser. 640, American Chemical Society, Washington, DC.

Environmental Protection Agency. 1998. Hazardous Waste management system; identification and listing of hazardous waste; land disposal restrictions, organobromine production waste. 40 CFR, Part 261.

Lai, D.Y., Woo, Y.-T., Argus, M.F., and Arcos, J.C. 1996. Cancer Risk Reduction Through Mechanism-Based Molecular Design of Chemicals. In *Designing Safer Chemicals,* ed. S.C. DeVito and R.L. Garrett, pp.62-73, ACS Symp. No. 640, American Chemical Society, Washington, DC

Lai, D.Y., and Woo, Y.-T. 1999. Structure-Activity Relationship and its Role in Cancer Risk Assessment. In *Genetic Toxicology and Cancer Risk Assessment,* ed. W.N. Choy, Marcel Dekker, N.Y., forthcoming.

Woo, Y.-T,. Lai, D.Y., Arcos,J.C., and Argu, M.F. 1984. *Chemical Induction of Cancer. Structural Bases and Biologic Mechanism. Vol. IIIB, Aliphatic and aromatic halogenated carcinogens.* Academic Press, Orlando Florida.

Woo, Y.-T., Lai, D.Y., Arcos, J.C., and Argus M.F. 1988. *Chemical Induction of Cancer. Structural Bases and Biologic Mechanism. Vol. III C, Natural, metal, fiber and macro-molecular carcinogens.* Academic Press, San Diego, CA.

Woo, Y.-T., Lai, D.Y., Argus M.F., and Arcos, J.C. 1995. Development of structure-activity relationship rules for predicting carcinogenic potential of chemicals. *Toxicol Lett* 79,219-228.

Woo, Y.-T., Lai, D.Y., Argus M.F., and Arcos, J.C. 1996. Carcinogenicity of organophosphorus pesticides/compounds: an analysis of their structure-activity relationships. *Environ Carcino Ecotoxicol Revs* C14:1-42.

Woo, Y.-T., Lai, D.Y., Argus M.F., and Arcos, J.C. 1997. Mechanism-based structure-activity relationship (SAR) analysis of carcinogenic potential of 30 NTP test chemicals. *Environ Carcino Ecotoxicol Revs* C15,139-160.

Woo, Y.-T., Lai, D.Y., Argus M.F., and Arcos, J.C. 1998. An Integrative Approach of Combining Mechanistically Complementary Short-term Predictive Tests as a Basis for Assessing the Carcinogenic Potential of Chemicals **_Environ Carcino Ecotoxicol Revs_ C16,101-122.**

Woo, Y.-T., and Lai, D.Y. 1999. Aromatic Amino and Nitro-Amino Compounds and their Halogenated Derivatives. In *Patty's Toxicology, Fifth Edition, Volume IV,* ed. E. Bingham, B. Cohrssen and C. Powell, John Wiley & Sons, New York, forthcoming.

Woo, Y.T., Lai, D.Y., McLain, J., Manibusan, M., Dellarco, V., and Cox, M. 1999. Structure-activity Relationship Analysis of Drinking Water Disinfection Byproducts for Potential Carcinogenicity. forthcoming.