Chemistry

85
Neural network based QSARs of chemical carcinogens derived from chemical safety database CAESAR

K. Tanabe, N. Ohmori, S. Ono and T. Suzuki[1]

Department of Management Information Science, Chiba Institute of Technology, Tsudanuma 2-17-1, Narashino-shi, Chiba 275-0016 and [1]Faculty of Economics, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan
kazutoshi.tanabe@it-chiba.ac.jp

The ability to assess the toxic potential of a chemical depends on the available information on the compound and/or its related compounds. Among chemicals currently in commerce, quite a few are ascertained on their toxicity including carcinogenicity and mutagenicity, and reliable data on chemical hazard assessment are very limited, especially for pharmaceutical chemicals. Therefore, an attempt on the basis of Quantitative Structure-Activity Relationship (QSAR) models for estimating the carcinogenicity has been performed (Romualdo 2003; Tanabe et al 2004, 2005). In this study a chemical safety database system named CAESAR (Computer-Aided Evaluation of Chemical Safety with QSAR), which consists of two databases, has been developed. One contains critically reviewed experimental toxicity data on selected chemical substances. Another contains toxicity data predicted on the basis of QSAR models for numerous, diverse chemicals in commerce. Experimental carcinogenicity data on more than 400 chemicals were collected from four sources including IARC (International Agency for Research on Cancer) and NTP (National Toxicity Program), and their reliabilities were ranked into nine categories based on their relative reliability of carcinogenicities by critically evaluating the six existing databases. CAESAR has a function of learning the relationship between chemical structure and carcinogenicity by using the artificial neural network (ANN) technology. For the ANN modeling, a three-layered network model to learn carcinogenic categories for diverse chemicals of 421 compounds was applied. Principal component analysis (PCA) was employed to select a smaller set of orthogonal descriptors from a pool of 70 molecular descriptors calculated with a CAChe Project-Leader (Fujitsu Ltd) from the 3D geometries of the compounds. The relationship between experimental carcinogenicity data and 25 descriptors (principal components) was analyzed with the ANN. After optimizing the architecture (25 input neurons for the 25 descriptors plus a bias, 20 intermediate neurons plus a bias, and one output layer neuron) and the conditions of the ANN, the performance of the derived model was assessed by applying the leave-one-out cross-validation test. The predicted values agreed well with the experimental ones, with the root mean squares error of 0.0069 and the squared correlation coefficient of 0.9994 for entire data set of 421 compounds. The results obtained in the prediction of the carcinogenicity by the model were excellent under the present uncertainties and difficulties of ranking of the experimental animal carcinogenic data. Application of CAESAR to other toxicity and endpoints such as mutagenicity and environmental endocrine disruption is being studied.

Acknowledgement: This work was supported by a Grant-in-Aid for Scientific Research (A) (14209022).

Romualdo, B. (2003) *Quantitative structure-activity relationship (QSAR) models of mutagens and carcinogens*. Boca Raton: CRC Press
Tanabe, K. et al (2004) *J. Pharm. Pharmacol*. **56** (Supplement): S58
Tanabe, K. et al (2005) *J. Ecotech. Res*. **11**: 111–116