The Utilization of the NTP-HTS Data in Chemical Toxicity Modeling

Hao Zhu[1], Ann Richard[2], Ivan Rusyn[3], and Alexander Tropsha[1]

[1]Laboratory for Molecular Modeling, School of Pharmacy, UNC-Chapel Hill; [2]National Center for Computational Toxicology, Office of Research & Development, US EPA; [3]Department of Environmental Sciences and Engineering, School of Public Health, UNC-Chapel Hill

INTRODUCTION

- A prominent goal of computational toxicology is to predict the potential adverse effects of chemical substances on living systems. Past structure-activity relationship approaches to prediction have relied exclusively upon identification of substructures or "toxic alerts". Newer approaches are exploring ways to incorporate richer descriptions of chemicals, in terms of their effects on fundamental biological targets and cells, into the prediction paradigm. The object is a prediction scheme based on a chemical profile consisting of both structural and biological properties.
- To reduce the high costs associated with animal toxicity testing and to incorporate greater efficiencies, the National Toxicology Program (NTP) at the National Institute of Environmental Health Sciences (NIEHS) recently initiated the High Throughput Screening (NTP-HTS) project (; Inglese et al 2006). This project aims to develop *in vitro* biological assays that would be predictive of *in vivo* chemical toxicity.
- In this study, the recent HTS results for known carcinogens were used as biological descriptors in Quantitative Structure Activity Relationship (QSAR)-based "disease-oriented" modeling. We show that the use of the HTS biological-content descriptors improves the predictive power of QSAR models of animal and human chemical carcinogenicity as compared to conventional models utilizing chemical descriptors only.

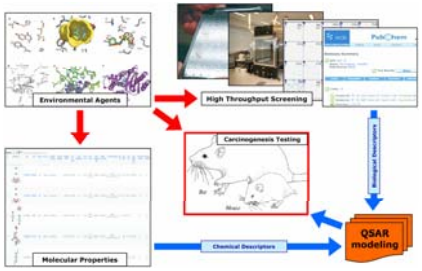

Figure 1. Combining chemical and biological descriptors in QSAR modeling of chemical carcinogenicity.

DATA

- The NTP-HTS data on 1,408 chemicals were obtained from PubChem (PUBCHEM) and the Distributed Structure-Searchable Toxicity (DSSTox) Database Network. After removing duplicates, inorganic, organo-metallic compounds and mixtures, the curated subset of the original NTP-HTS dataset used in our studies included 1,289 unique chemical agents (**Table 1**).
- The cell lines used for screening in NTP-HTS include: BJ (human foreskin fibroblast), HEK293 (transformed human embryonic kidney cell line), HepG2 (human hepatoma), Jurkat (clone E6-1, human acute T cell leukemia), MRC-5 (human lung fibroblast) and SK-N-SH (human neuroblastoma).
- The rodent carcinogenicity data were obtained from the DSSTox version of the Carcinogenic Potency Database (CPDBAS) (Gold et al 1997). The dataset used in this study included a total of 270 compounds: 178 were active and 92 inactive. The rodent carcinogenicity data for these compounds in four different species are shown in **Table 2**.

Table 1. NTP-HTS screening outcome for 1,289 Compounds

Classifications	BJ	HEK293	HepG2	Jurkat	MRC-5	SK-N-SH	All tests
Actives	42	63	41	121	37	74	140
Inconclusives	44	79	47	89	44	54	90
Inactives	1,203	1,147	1,201	1,079	1,208	1,161	1,059

Table 2. Rodent Carcinogenicity Results for 270 Compounds from NTP-HTS Database

	Male Rats	Female Rats	Male Mice	Female Mice
Active	122	112	124	134
Inactive	130	135	125	113
Total	252	247	249	247

QSAR METHODOLOGY

- Chemical carcinogenicity data were analyzed using the predictive QSAR modeling framework (**Figure 2**) developed in our laboratory. The workflow incorporates modules for combinatorial QSAR model development (i.e., using all possible binary combinations of available descriptor sets and statistical data modeling techniques), rigorous model validation, and virtual screening of available chemical databases to prioritize chemicals for toxicity testing. Chemical toxicity models were developed using chemical descriptors only, or using chemical descriptors combined with NTP-HTS biological descriptors.
- Most computations were carried out using kNN QSAR modeling method developed in this laboratory (Zheng et al., 2000). Chemical descriptors reflecting molecular connectivity, shape, and charge distribution were calculated with the MolConnZ software.

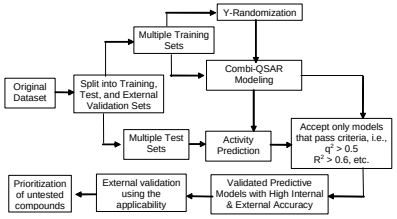

Figure 2. Flowchart of predictive toxicology framework based on validated combi-QSAR models.

RESULTS & DISCUSSION

- The relationship between HTS activity and rodent carcinogenicity of 270 compounds is shown in **Table 3**. As indicated, HTS active compounds have 81% chance to be rodent carcinogens. On the contrary, an HTS inactive compound is not necessarily expected to be non-carcinogenic (only 35% of HTS inactives are CPDB inactives as well). Thus, the HTS data alone are insufficient to forecast the *in vivo* carcinogenicity.

Table 3. The relationship between HTS activity and rodent carcinogenicity of 270 compounds.

	HTS actives	HTS inconclusives	HTS inactives
CPDB actives	30	12	136
CPDB Inactives	7	11	74
Correlation	81%	-	35%

- The same dataset was analyzed with the conventional QSAR approach using chemical descriptors of 270 compounds only. Using the computational workflow of Fig. 2 we identified nine rigorously validated *k*NN QSAR models that had high prediction accuracy for both training and test sets (q^2 *and* R^2, respectively) greater than 0.7.
- We then examined whether the use of *in vitro* bioassay (HTS) data as biological fingerprint descriptors could improve the prediction accuracy of QSAR models. We found that when using the combined chemical and seven HTS biological descriptors the number of *k*NN QSAR models that satisfy the $q^2/R^2 > 0.7$ cutoff increased to 34.
- Not only the number of successful kNN QSAR models increased but the external prediction accuracy of the models relying on combined chemical and biological descriptors improved as well. **Figure 3** compares the prediction accuracy of conventional QSAR models vs. those generated with the combined descriptors for 46 compounds in the external evaluation set. The data indicate that the average accuracy of the models using the combined descriptors is 17% higher.

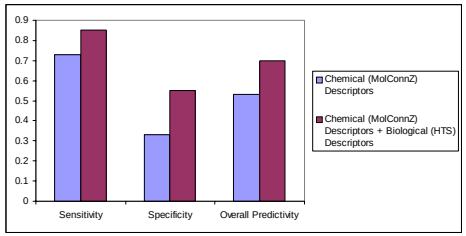

Figure 3. Comparison of the results from *k*NN QSAR models using two types of descriptors.

- HTS results generated in six different cell lines were given equal weights in assigning the overall HTS activity of a compound. However, we examined the differential significance of each cell line in contributing to QSAR models. **Figure 4** shows that the frequency of use of the seven HTS descriptors in the 34 *k*NN QSAR models is not the same. This distribution suggests that the choice of cell lines could significantly influence the significance of HTS screening in predicting chemical carcinogenicity *in vivo*.

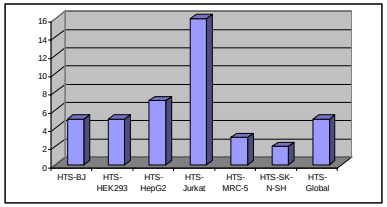

Figure 4. Seven HTS descriptors with their frequency of use in the 34 *k*NN QSAR models.

CONCLUSIONS

- We have examined the usefulness of NTP-HTS data for predicting chemical toxicity *in vivo*. We have shown that the current HTS results, from assays not deliberately chosen for carcinogenicity relevance, have limited predictive power (Table 3) by themselves. However, using HTS results as biological fingerprint descriptors significantly improved the overall prediction accuracy of QSAR models as compared to those based on chemical descriptors alone (Figure 3).
- This observation suggests that *in vitro* bioassay results add new information content useful in prediction and may be ultimately helpful in at least partially replacing *in vivo* toxicity testing. This is most likely to be the case if the battery of HTS assays is expanded to include cell lines and processes with known relevance to carcinogenicity.

REFERENCES

- DSSTox: http://www.epa.gov/ncct/dsstox/index.html
- Gold, L. S., Slone, T. H., Ames, B. N., Summary of Carcinogenic Potency Database by Chemical. In: Gold, L. S., Zeiger, E. (eds), Handbook of Carcinogenic Potency and Genotoxicity Databases, 1997 CRC Press, Boca Raton , Fla, pp. 624-657.
- NTP High Throughput Screening Program, http://ntp-server.niehs.nih.gov/index.cfm?objectid=05F80E15-F1F6-975E-77DDEDBDF38941CD
- Inglese, J., Auld, D. S., Jadhav, A., Johnson, R. L., Simeonov, A., Yasgar, A., Zheng, W., Austin, C. P. Quantitative high-throughput screening: A titration-based approach that efficiently identifies biological activities in large chemical libraries. PNAS 2006, 103: 11473-11478.
- PUBCHEM: http://pubchem.ncbi.nlm.nih.gov/
- Zheng, W., Tropsha, A. A novel variable selection QSAR approach based on the K-nearest neighbor principle. J. Chem. Inf. Comput. Sci, 2000; 40; 185-194.