OAK RIDGE
NATIONAL LABORATORY
MANAGED BY UT-BATTELLE
FOR THE DEPARTMENT OF ENERGY

NFE-06-00020

Environmental Sciences

CRADA Final Report
For
CRADA Number
NFE-06-00020

LiverTox: Advanced QSAR and Toxicogeomic Software for Hepatotoxicity Prediction

Po-Yun Lu
Oak Ridge National Laboratory

Dr. Katherine Yuracko
YAHSGS, LLC

Prepared by
Oak Ridge National Laboratory
Oak Ridge, TN 37831
managed by
UT-BATTELLE, LLC
for the
U.S. Department of Energy
under contract DE-AC05-00OR22725

Approved for

Unlimited Release

UT-BATTELLE
ORNL-27 (4-00)

LiverTox: An Advanced Toxicogenomic System for Hepatotoxicity Prediction

CRADA Final Report

CRADA No. NFE-06-00020 UT-Battelle LLC and YAHSGS LLC

Sponsored by National Institute of Health's National Institute of Environmental Health Sciences
SBIR/STTR Grant No. 5R42ES01321-04

Collaborator: YAHSGS LLC PI: Kathy Yuracko Address: 3100 George Washington Way, Suite 103, Richland, WA 99352 Phone: (509) 539-7147

Research Institution: Oak Ridge National Laboratory PI: Po-Yung Lu Address: 1060 Commerce Park, Oak Ridge, TN 37830 Phone: (865) 574-7587

Abstract

YAHSGS LLC and Oak Ridge National Laboratory (ORNL) established a CRADA in an attempt to develop a predictive system using a pre-existing ORNL computational neural network and wavelets format. This was in the interest of addressing national needs for toxicity prediction system to help overcome the significant drain of resources (money and time) being directed toward developing chemical agents for commerce. The research project has been supported through an STTR mechanism and funded by the National Institute of Environmental Health Sciences beginning Phase I in 2004 (CRADA No. ORNL-04-0688) and extending Phase II through 2007 (ORNL NFE-06-00020). To attempt the research objectives and aims outlined under this CRADA, state-of-the-art computational neural network and wavelet methods were used in an effort to design a predictive toxicity system that used two independent areas on which to base the system's predictions. These two areas were quantitative structure-activity relationships and gene-expression data obtained from microarrays. A third area, using the new Massively Parallel Signature Sequencing (MPSS) technology to assess gene expression, also was attempted but had to be dropped because the company holding the rights to this promising MPSS technology went out of business. A research-scale predictive toxicity database system called Multi-Intelligent System for Toxicogenomic Applications (MISTA) was developed and its feasibility for use as a predictor of toxicological activity was tested. The fundamental focus of the CRADA was an attempt and effort to operate the MISTA database using the ORNL neural network. This effort indicated the potential that such a fully developed system might be used to assist in predicting such biological endpoints as hepatotoxcity and neurotoxicity. The MISTA/LiverTox approach if eventually fully developed might also be useful for automatic processing of microarray data to predict modes of action. A technical paper describing the methods and technology used in the CRADA has been published. This paper was entitled "A Toxicity Evaluation and Predictive System Based on Neural Networks and Wavelets" and appeared in an American Chemical Society peer-reviewed publication this year (*J. Chem. Inf. Model.* 47: 676685, 2007). A patent application was filed but later abandoned.

Background

The MISTA/LiverTox system sought to be developed is based on the database known as Multi-Intelligent Systems for Toxicogenomic Applications (MISTA), jointly developed by ORNL and YAHSGS employees. Its potential feasibility during Phase I had been earlier described in CRADA No. ORNL-04-0688 with YAHSGS LLC. MISTA/LiverTox (LiverTox) was supported by a biological database developed during Phase I and attempts were to be during Phase II of the project to further develop this effort. This database was intended if the system was fully developed to be used to train the battery of computational neural networks (CNNs) which needed to be developed to enhance LiverTox predictions. Extrapolating from Phase I research results and taking into account the planned Phase II developments described herein, a Phase II research objective for LiverTox was suppose to be an effort to hopefully achieve a predictive accuracy greater than 95% on average, with no endpoints below the 90% confidence level. However, these accuracies were not obtained and could not be reached by the conclusion of this effort. As Phase II progressed, LiverTox attempted to use the Interagency Coordinating Committee on Validation of Alternative Methods (ICCVAM) protocols. During the project's last phase, YAHSGS was solely responsible for developing and adapting the system to be more user-friendly using

Graphic User Interface (GUI) technology to be developed by YAHSGS and arrive at an integrated predictive toxicity platform that could be acclimated for commercial market applications. However, this was not achieved.

Statement of Objectives-Specific Aims and Tasks

A. Phase II Aims (12/20/2005 to 3/31/2007)

Aim 1. Adapt MISTA/LiverTox to potentially achieve efficient and accurate data retrieval, exchange, and filing for existing and forthcoming LiverTox modules.

Aim 2. Attempt to enhance prediction accuracy of integrated QSAR and microarray modules through enhanced microarray training sets, and verify ability to make integrated predictions.

Aim 3. Attempt to add, train, and potentially verify functionality of new additional separate and parallel MISTA/LiverTox modules using MPSS, metabonomic/metabolomic, and proteomic datasets.

Aim 4. Attempt to integrate all MISTA/LiverTox modules (QSAR, microarray, MPSS, proteomic, and metabonomic/metabolomic) and verify ability to make liver-toxicity predictions using combined knowledge base.

Aim 5. Try to validate MISTA/LiverTox using protocols based on ICCVAM and challenge chemical groups that the National Toxicogenomics Center was to provide.

B. Phase II Extension Period Tasks (4/1/2007 to 12/31/2007)

Task 1. It was agreed that beginning April 1, 2007, all further development of computational techniques for LiverTox was to be performed by YAHSGS employees or consultants. YAHSGS had access to project information including algorithms, data files, formulas, codes, programs, and designs related to LiverTox.

Task 2. If requested, ORNLwould attempt to consult on the design of MISTA/LiverTox computation effort.

Task 3. ORNL would attempt to interface with ICCVAM in support of an effort hopefully toobtain ICCVAM validation of LiverTox.

Task 4. Morey Parang, a University of Tennessee employee and a YAHSGS LLC consultant was tasked by YAHGS to attempt to integrate MISTA/ LiverTox to produce a toxicity prediction tool for commercial deployment.

Task 5. Final Report Preparation.

Technical Project Accomplishments

A. Phase II (12/20/2005 to 3/31/2007)

Aim 1: a. Macro software was used to aid in data entry to be integrated into the web-based data management system.
b. Attempt to verifythe functionality of data retrieval software and compatibility with MISTA/LiverTox using Phase I microarray modules and hepatotoxicity.
c. It was hoped that enhancements might make it possible for QSAR information to be more readily extracted and used in molecular mechanics, molecular dynamics, and normal-mode simulations as well as semiempirical chemical data than previous capabilities allowed. If so, MISTA might be complimented by an effort to develop the use of computational chemistry to predict biological events, such as hepatotoxicity, genotoxicity, neurotoxicity, partition efficients, metabolism, and model of action. See Attachment 2 ("A Toxicity Evaluation and Predictive System Based on Neural Networks and Wavelets,"

P.L. Piotrowski, B.G. Sumpter , H.V. Malling, J.S. Wassom, P.Y. Lu, R.A. Brothers, G.A. Sega, S.A. Martin, and M. Parang, *J. Chem. Inf. Model.* **47**:676-685, 2007).

Aim 2: A liver microarray toxicity database called MISTA was developed from open peer-reviewed literature including chemical name, CAS Reg. No., gene class, symbol, physio/biochemical category, gene bank number, OMIM number, animal and dose information, and expression results. Field entries for microarray type were checked, and gene symbols/names were verified against current official gene symbol/name. The database has a total of 15825 data records and 3159 genes (distinct gene symbol verified by MISTA: 2653 in 212 distinct gene classes). More information can be found in Other Accomplishments 1. (MISTA Web Application: Hepatotoxicity Database).

Aim 3: This aim had to be changed because the MPSS system could not achieve the level of implementation expected, thus the availability of data was insufficientto be used. Also, the company (Lynx Therapeutics, Inc.) that developed MPSS was dissolved. For this reason, we were not able to pursue this objective further.

Aim 4: a. The predictive capability of the system was attempted to be tested using the MISTA database and other data as training sets. Improved accuracy of the pre-existing system was obtained but only in a highly controlled research setting. Progress was made toward integration of modules but was not achieved due to changes made by YAHGS in project direction and personnel changes.
An attempt was planned to add a new descriptor calculation to the pre-existing QSAR program. It was intended to approximate the aromatic character of a compound using the HOMA index. The QSAR program was intended to calculate 5 descriptors: molecular weight, moments of inertia, topological indices, wavelet transform of the 3-D structural information, and aromaticity. If achieved, it was hoped the user would be allowed to choose to use some or all of these descriptors as input into the prediction program.
In order to train the prediction program a set of known data was needed. The MISTA

database was used to extract data and create a training set. A molecular file (.mol) was input into the QSAR program. For new molecules, this file can be generated using ChemSketch, or similar software. The molecular files were input into the QSAR program and the descriptors were calculated. This descriptor file was then to be used as input into the prediction program. The output from the QSAR program and the known results were to be put into the prediction program for training.

The predictive power of the MISTA system was tested in a laboratory setting and the ability of the MISTA neural network-base approach could be demonstrated but only in a highly controlled research setting with highly trained ORNL researchers, and by the CRADA end YAHSGS had been unable to develop a user friendly prediction program.

Aim 5: No expert panels were convened by YAHGS to challenge the data sets developed due to reasons stated in 4d above.

B. Phase II Extension Period (4/1/2007 to 12/31/2007)

Task 1:

All pertinent information concerned with development of the MISTA/LiverTox system was compiled onto two CDs and transferred to Morey Parang, the YAHSGS LLCconsultant, who took over responsibility for development of the MISTA/LiverTox module, and Dr. Yuracko on April 26, 2007 undertook the responsibility completion of Phase II. . Copies of the CDs were given to Mr. Parang.

Task 2:

Informal discussions were held with Mr. Parang for the purpose of providing user interface input and output screens. A document on this subject was prepared and the original draft version was sent to Morey Parang on September 24, 2007 and the subsequent revision on December 5, 2007, adding the interactive diagram of MISTA/LiverTox and the future subscribers/users.

Task 3:

Several phone discussions were held with Dr. Raymond Tice, Deputy Director, ICCVAM, NIEHS, about the potential use of MISTA/LiverTox as an alternative toxicity testing system to replace *in vivo* whole animal toxicity assays according to the ICCVAM Authorization Act of 2000 (Public Law 106-545). In addition to these telephone discussions, a meeting was held May 24, 2007, with. Tice at NIEHS to review the current status and usefulness of LiverTox as published in the paper by Piotrowski et al. 2007. Dr. Tice concluded hat the MISTA/LiverTox system showed promise as a predictive toxicology system. Before being considered for submission for ICCVAM review and evaluation, he concluded that it needs additional work focusing on the system's applicability as an alternate test method, as described above, or one that aids and abets other alternate methods.

Task 4:

YAHSGS began efforts to integrate the MISTA/ LiverTox in an effort to develop a user-friendly, commercially deployable product.

Task 5: A final report was to be submitted.

Accomplishments 2

Patent application

Based on the research findings detailed in Piotrowski, et al. 2007 (Attachment 2), a patent application was filed entitled “LiverTox - A Toxicity Evaluation and Predictive System Based on Neural Networks and Wavelets," on July 17, 2007; ORNL ID: 1912. The application has been abandoned.

Publication

A referred journal article entitled “A Toxicity Evaluation and Predictive System Based on Neural Networks and Wavelets” by P.L. Piotrowski, B.G. Sumpter , H.V. Malling, J.S. Wassom, P.Y. Lu, R.A. Brothers, G.A. Sega, S.A. Martin, and M. Parang was published in *J. Chem. Inf. Model*. **47**:676-685, 2007. (Attachment 2)

Presentations

1 Malling, H. M., Wassom, J. S., Sega, G., Sankaranarayanan, K, and Lu, P.-Y., “Predicting Idiosyncratic Toxicity Responses of Xenobiotics in Rodent Livers.” Poster presentation. 37th Annual Meeting, Environmental Mutagen Society, Sept. 16-20, 2006. British Columbia, Canada.

2 Wassom, J. S., Malling, H., Sega, G., Sankaranarayanan, K., Brothers, R., Parang, M., Lu, P.-Y., Sumpter, B., Piotrowski, P., Martin, S., Martin, K., Goldberg, J. and Edmond, A., “An *In Silico* Predictive System for Evaluating Induced Liver Toxicity in Rats and Extrapolation to Human Liver Toxicity.” Poster presentation. 37th Annual Meeting, Environmental Mutagen Society, Sept. 16-20, 2006. British Columbia, Canada.

3 Piotrowski, P., Sumpter, B., Lu, P.-Y., Martin, S., Parang, M., and Wassom, J.S. “A Predictive System for Computational Molecular Toxicology Based on Neural Network.” Poster presentation. 46th Annual Meeting of the Society of Toxicology, March 25-29, 2007, Charlotte, N.C.

Benefits to the Funding DOE Office’s Mission

The work and research associated with this CRADA agreement were supported by the National Institute of Environmental Health Sciences (NIEHS). By way of this CRADA, ORNL and YAHSGS LLC intended to enhance the mission of DOE’s Office of Science by collaborating to produce a system that hopefully would complement the management of huge volumes of genomic-related high-throughput data generated by or stimulated by the Human Genome Project and the Genomics:GTL Program. Additionally, the work was intended to provide DOE with a means by which the agency can assess the toxicity of chemical agents.

Technical Discussion of Work Performed by All Parties

Tasks in which ORNL participated:

1 Attempted feasibility of applying ORNL neural network and Wavelet software to evaluate chemical effects on the genomes of single or multiple organisms, focusing on genes of significance to liver toxicity and chemicals from classes of concern to human health.

2 Effort to develop and adapt algorithms for toxicity assessment of chemical actions and for understanding the significance of these interactions and how such toxicological events lead to specific disease states.

3 Effort to integrate YAHSGS advanced structure-toxicity modules to simultaneously predict chemical toxicity independent of microarray data, using a prototype to demonstrate the system's applicability by challenging it with microarray test results.

4 Effort to develop a web-based structure to store and use project data, with features that promote efficient receipt and exchange of information from linked researchers.

5 Effort to develop and maintain toxicity data for the MISTA/LiverTox database.

6 Interacted with ICCVAM/NIEHS as to whether the MISTA/LiverTox approach might be or with additional research effort become suitable to submit for a formal evaluation as an alternative test method for toxicity testing.

Tasks in which YAHSGS participated:

1 Led project's overall technical and administrative management.

2 Evaluated toxicity data for the MISTA/LiverTox hepatotoxicity database.

3 Solely responsible for the development of a user-friendly pilot commercial product of MISTA/LiverTox.

4 Responsible for exploring marketing possibilities and strategies for further development of MISTA/LiverTox after attempted completion of Phase II efforts.

Commercialization Possibilities

The MISTA platform and database with further additional research effort and hopeful results, could potentially be developed further and, when ready, considered by ICCVAM as an alternate testing device to enhance clinical screening processes and to reduce whole animal use in toxicity testing.

Conclusions

The collaboration between ORNL and YAHSGS LLC during Phase II of the CRADA project made some progress in an attempt to develop a research-scale MISTA/LiverTox system. This system sought to use a pre-existing ORNL neural network and wavelet-based format with only some limited success in a highly controlled research setting. Aspects of this system were submitted for patent approval on July 17, 2007, but the patent application has been abandoned. YAHSGS was responsible for the development of commercially deployable versions of the LiverTox system. As of the date of this Report, no such versions have been created and/or developed.

A referred journal article entitled “A Toxicity Evaluation and Predictive System Based on Neural Networks and Wavelets” by P.L. Piotrowski, B.G. Sumpter , H.V. Malling, J.S. Wassom, P.Y. Lu, R.A. Brothers, G.A. Sega, S.A. Martin, and M. Parang was published in *J. Chem. Inf. Model*. **47**:676-685, 2007. (Attachment 2)