Final Report
of the Air Force Project

Covering research period 7/15/05 to 7/14/08

Agency No: DOD/FA9550-05-1-0456
U of M No: 1613-189-6286
NRRI Technical Report No: NRRI/TR-2008/18

Predicting Chemical Toxicity from Proteomics and Computational Chemistry

July 30, 2008

Submitted by:

Subhash C. Basak, Ph.D.
Principal Investigator
Natural Resources Research Institute
University of Minnesota, Duluth
5013 Miller Trunk Highway
Duluth, MN 55811
Tel: (218)720-4230
Fax: (218)720-4328
Email: sbasak@nrri.umn.edu

I. Table of Contents

II. Objectives

The primary objectives of this project were:

a) Development of novel biodescriptors to characterize proteomics patterns (maps)
b) Use of these new biodescriptors and those formulated under the previous AFOSR grant to predict chemical toxicity from toxicoproteomics data
c) Use of chemodescriptors to develop hierarchical quantitative structure-toxicity relationship (HiQSTR) models to predict the toxicity of chemicals from molecular structure
d) Comparative studies of biodescriptors vis-à-vis chemodescriptors in predictive toxicology
e) Development of integrated QSTR models using the combined set of chemodescriptors and biodescriptors
f) Study and modeling of chemical hormetic potency using existing data from the NCI/NIH Developmental Therapeutics Program (DTP)

A major part of the project resources were focused on developing novel biodescriptors and the use of already formulated biodescriptors in predicting chemical toxicity. The other important objectives were to investigate the utility of chemodescriptor-based HiQSTR models in predicting the toxicity of halocarbons and peroxisome proliferators, as well as chemisorption of JP8 chemicals to membrane coated fibers. The biodescriptor portion of the project also explored the utility of combining biodescriptors and chemodescriptors in predictive toxicology. Finally, some limited resources were directed to acquiring and modeling the phenomenon of hormesis, as indicated in cell culture data from the NCI/NIH Developmental Therapeutics Program. These objectives were accomplished in terms of the following specific tasks:

1. Databases of chemical properties, activities, and toxicities from open literature and sources within the US Air Force Research Laboratory were established
2. Additional development of abundance-graph biodescriptors
3. Development of information-theoretic invariants from graphs of proteomic maps
4. Development of novel biodescriptors from spectrum-like representation of proteomic patterns
5. Selection of protein spot biodescriptors using robust statistical methods
6. Compared proteomic maps using the various classes of biodescriptors
7. Utilized biodescriptors in predicting toxicity of chemicals
8. Developed and applied chemodescriptors in HiQSTR analysis
9. Formulated integrated QSTRs to predict the toxicity of chemicals
10. Formulated HiQSTRs for membrane coated fiber (MCF) data using calculated chemodescriptors
11. Applied novel similarity and tailored similarity measures to cluster JP-8 chemicals for laboratory testing
12. Compared chemodescriptors and biodescriptors in predicting the toxicity of chemicals
13. Developed a structure and activity database for chemicals demonstrating hormetic activity in the NCI Yeast Anticancer Drug Screen program
14. Conducted exploratory modeling of hormetic potency for the large data set from the NCI Yeast Anticancer Drug Screen program

III. Status of Effort

Due to budgetary constraints, research on modeling proteomic data was limited to data sets already available to the research team: the four peroxisome proliferators, fourteen halocarbons, and data on human keratinocyte exposure to JP-8. In terms of this available data, the peroxisome proliferator data set has been used exhaustively in the formulation of novel biodescriptors. This set has served as the primary set for fundamental biodescriptor development, and has been used in initial testing of all biodescriptors derived from proteomic maps. The research team has done extensive modeling of halocarbon toxicity using the HiQSAR (hierarchical QSAR) approach, including high-level quantum chemical calculations provided by Dr. K. Balasubramanian. This set has also been modeled with the current set of biodescriptors and has served as the test set for I-QSAR (integrative QSAR) development —modeling with the combined set of chemodescriptors and biodescriptors. Toxicity models based on chemodescriptors are promising. Finally, the keratinocyte data has been modeled using a comprehensive set of biodescriptors. Unfortunately, given the nature of the cellular treatment, exposure to the complex JP-8 mixture, and the lack of functional endpoints, use of chemodescriptors to model changes in cellular function was impossible.

This work on biodescriptors, and the combination of chemodescriptors and biodescriptors to model chemical toxicity, has led to a large number of publications during the term of this project. As can be seen in *Section VI, Publications*. This project's biodescriptor research has resulted in the authorship of fourteen additional peer-reviewed publications on the calculation and application of biodescriptors in modeling the perturbation of the cellular proteome. In fact, this work funded primarily by AFOSR has opened a new field of research—mathematical proteomics. The NRRI research team's efforts as leaders in the emerging field of mathematical proteomics and biodescriptor development was recently recognized by the editors of the Thomson Scientific journal *Current Opinions in Drug Discovery and Development* when they invited Dr. Basak to write a review on the current state and future directions in the field of mathematical biodescriptor development and use [Basak and Gute, 2008].

Additionally, this work in mathematical proteomics inspired Basak and colleagues to develop biodescriptors to characterize the primary structure of DNA and RNA using mathematical invariants. While some work on graphical representations of DNA primary structure had already been done, the application of mathematical invariant techniques by Basak and colleagues opened the flood gates on the development of novel invariants as DNA and RNA sequence descriptors. This began as collaborative work with an Indian scientist, Dr. Ashesh Nandy, and the publication of several papers on methods to characterize and compare DNA sequences. However, this has led to explosive growth in this emerging field in which hundreds of studies have now been published. One of the most significant results of this work led to a publication in the *Journal of Chemical Information and Modeling* showing the method's ability to characterize and compare the neuraminidase gene sequence of H5N1 avian flu resulting in observations that could help to identify future strains of the virus that will be highly pathogenic to humans [Nandy, Basak and Gute, 2007]. This work was funded in part by this AFOSR project along with a grant from the University of Minnesota's Consortium for Bioinformatics and Computational Biology.

The additional work on QSAR modeling and method development (including the work on I-QSAR methods) has successfully introduced novel statistical approaches to the QSAR field—namely the concept of naïve (or over-fitted) Q^2 as compared to true Q^2. The field of QSAR research has long been subject to misconceptions and misinterpretations of proper statistical methods. Thanks in large part to the efforts of Dr. Hawkins in collaboration with Dr. Basak, the two have been spreading the message of the dangers of over-fitting and the misconceptions about Q^2 as a useful measure of modeling "success" that have been held by certain well-known individuals in the field of QSAR research and development. Efforts in QSAR modeling during this project have also included studies related to developing statistically robust models and the use of proper descriptor thinning techniques in combination with model evaluation to ensure that over-fitting or over zealous trimming of descriptors does not lead to erroneous conclusions regarding the overall usefulness and reliability of the model. These techniques have been applied and validated using the a congeneric set of halocarbons, some more heterogeneous data sets (including modeling the vapor pressure for a large, diverse set of chemicals), and in successfully modeling and predicting pharmacokinetics.

Finally, the efforts in hormesis modeling were complicated by the great diversity of the chemical data set tested by NCI and the lack of emphasis on the hormetic nature of these compounds (researchers generating the data were far more interested in characterizing LOAELs [lowest observed adverse effect levels] and LC_{50} [lethal concentration in 50% of test subjects], than in carefully characterizing hormetic potency and dose range). Working with Dr. Edward Calabrese's group at the University of Massachusetts – Amherst, chemicals within the set were identified as having strong or weak hormetic potency for the purposes of modeling. While there was no evidence of a hormetic dose range for some of the chemicals within the data set, the data were not sufficient to state conclusively that these compounds were not hormetic. It is unfortunate that this area of toxicology research has been largely ignored by the scientific community, and therefore little actual testing for hormetic potency has been conducted. As a result, insufficient testing has been conducted at low dose ranges for many compounds, leaving their hormetic potency (or lack of hormetic activity) a matter of speculation. Given the large scope of the task — developing the hormesis data set from the NCI web site and modeling hormetic potency — the research team was only able to conduct preliminary modeling. While the results of the modeling were promising, the significant structural diversity in the data set will require significant future work to examine more narrowly defined structural classes for more precise modeling of the hormetic phenomenon.

IV. Accomplishments/New Findings

The major effort of this project was focused on further development and use of biodescriptors for proteomics maps. This effort was roughly divided into two major categories: a) development and validation of novel biodescriptors and b) combining chemodescriptors and biodesciptors for biologically-enhanced QSTR modeling. In the category of biodescriptors, three approaches have been pursued: 1) the creation of numerical invariants (or vectors of invariants) derived from embedded graphs associated with proteomics maps (2-DE gels); 2) the

development of vectors derived from projections of three dimensions (protein mass, charge, and abundance) onto three planes, the (x,y), (y,z) and (x,z) planes, and 3) the development of information theoretic indices for proteomics maps based on partitioning the electrophoretic gel into *nxn* cells.

Results for the numerical invariants based on proteomics maps from liver tissue from rats exposed to peroxisome proliferators (perfluorooctanoic acid, perfluorodecanoic acid, clofibrate, and diethylhexyl phthalate) show that the leading eigenvalue of the D/D matrix derived from embedded graphs shows reasonable power to discriminate among maps derived from mechanistically and structurally similar chemicals. Several new studies on numerical invariants as biodescriptors have been published as a result of continued research efforts in this area [Randic et al, *J. Proteome Res.*, 2005; Balasubramanian et al, *J. Proteome Res.*, 2006; Randic et al, *J. Chem. Inf. Model.*, 2006].

The approach derived from 3D projections led to the creation of a vector space. Euclidean distance in such a space can be used as a measure of the similarity/dissimilarity of proteomics maps. This approach has been shown to cluster peroxisome proliferators and halocarbons reasonably well [Vracko et al, *J. Chem. Inf. Model.*, 2006; Bielinska-Waz et al, *Eur. Phys. J. B.*, 2006; Bielinska-Waz et al, *Symmetry, Spectroscopy, and SCHUR*, 2006].

The information theoretic approach developed by Basak et al partitions the entire (*x,y*) plane defined by protein mass and charge into a certain number of cells, *nxn*, where proteins in each cell are considered equivalent. Shannon's relation was then used to compute the complexity of the entire map. Results for the four peroxisome proliferators, *viz.*, PFOA, PFDA, clofibrate, and DEHP, show that this approach clusters the first three highly-fluorinated and mechanistically similar chemicals together, while putting DEHP in a category by itself. The efficacy of this novel approach has been further validated in several studies [Basak et al, *WSEAS*, 2005; Basak et al, *Computation in Modern Science and Engineering*, 2007; Basak and Gute, *Curr. Opin. Drug Disc. Dev.*, 2008; Basak et al, *Principles and Practice of Mixture Toxicity*, submitted].

In addition to the work on proteomics biodescriptors, effort was also devoted to developing biodescriptors (mathematical invariants) for characterizing DNA and RNA sequences. This work, originally pioneered by Basak, Nandy and Randic, has led to a number of recent publications [Randic et al, *Chem. Phys. Lett.*, 2005; Nandy et al, *ARKIVOC*, 2006]. These biodescriptors have been applied to study genetic variability that could be key to differentiating highly-virulent strains of common viruses (such as H5N1, the avian influenza virus) from those strains that do not show virulence in humans (or other species) [Nandy et al, *J. Chem. Inf. Model.*, 2007].

Finally, it must be noted that little of any significance has resulted from the work on hormesis. While the NRRI team worked closely with Dr. Calabrese and his colleagues (primarily Mark Nascarella), there were serious shortcomings to the data available on the NCI web site. As was mentioned previously, due to the poor quality of the available data and the broad chemical nature of the data set, modeling efforts were not satisfactory. Further study in modeling hormesis will require significant effort in cleaning the available data and characterizing the chemicals into better defined structural classes.

V. Personnel Supported

Subhash C. Basak, Principal Investigator (NRRI, University of Minnesota Duluth, Duluth, MN)

Douglas Hawkins, Co-principal Investigator (Professor—School of Statistics, University of Minnesota, St. Paul, MN)

Brian Gute, Research Fellow (NRRI, University of Minnesota Duluth, Duluth, MN)

Denise Mills, Research Coordinator (NRRI, University of Minnesota Duluth, Duluth, MN)

Varsha Kodali, Graduate Research Assistant (Department of Computer Science, University of Minnesota Duluth, Duluth, MN)

Terrence Neumann, Graduate Research Assistant (Department of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth, MN)

VI. Publications

The following 47 peer-reviewed papers and book chapters, which are currently either published, in press, or submitted, report results of research carried out between July 15, 2005 and July 14, 2008.

2005

Canonical labeling of proteome maps, M Randić, N Lerš, D Vukičević, D Plavšić, BD Gute and SC Basak, *J. Proteome Res.*, **4**, 1347–1352 (2005).

Four-color map representation of DNA or RNA sequences and their numerical characterization, M Randić, N Lerš, D Plavšić, SC Basak and AT Balaban, *Chem. Phys. Lett.*, 407, 205–208 (2005).

Information-theoretic biodescriptors for proteomics maps: Development and applications in predictive toxicology, SC Basak, BD Gute and FA Witzmann, *WSEAS Transactions on Information Science and Applications (Proceedings of the 9th WSEAS International Conference on Computers)*, **7**, 996–1001 (2005).

My tortuous journey from biochemistry to mathematical chemistry, SC Basak, in *Proceedings of the 50th Anniversary Symposium of the Department of Biochemistry, University of Calcutta, Kolkata, India.*

Prediction of partitioning properties for environmental pollutants using mathematical structural descriptors, SC Basak and D Mills, *ARKIVOC*, **2005**, 60–76 (2005), www.arkat-usa.org.

Predicting permeability of antimycotics from calculated chemodescriptors: A hierarchical QSAR approach, SC Basak and D Mills, *WSEAS Transactions on Information Science and Applications (Proceedings of the 9th WSEAS International Conference on Computers)*, **7**, 954–957 (2005).

Quantitative structure-activity relationship modeling of insect juvenile hormone activity of 2,4-dienoates using computed molecular descriptors, SC Basak, R Natarajan, D Mills, DM Hawkins and J Kraker, *SAR QSAR Environ. Res.*, **16**, 581–606 (2005).

Structure-activity relationships for mosquito repellent aminoamides using the hierarchical QSAR method based on calculated molecular descriptors, SC Basak, N Ramanathan and D Mills, *WSEAS Transactions on Information Science and Applications (Proceedings of the 9th WSEAS International Conference on Computers)*, **7**, 958–963 (2005).

2006

Combining chemodescriptors and biodescriptors in quantitative structure-activity relationship modeling, DM Hawkins, SC Basak, J Kraker, KT Geiss and FA Witzmann, *J. Chem. Inf. Model.*, **46**, 9–16 (2006).

Complex graph matrix representations and characterizations of proteomic maps and chemically induced changes to proteomes, K Balasubramanian, K Khokhani and SC Basak, *Proteome Res.*, **5**, 1133–1142 (2006).

Complexity of chemical graphs in terms of size, branching, and cyclicity, AT Balaban, D Mills, V Kodali and SC Basak, *SAR QSAR Environ. Res.*, **17**, 429–450 (2006).

Fourth Indo-U.S. Workshop on Mathematical Chemistry, January 8–12, 2005, Maharashtra, India (Editorial), DK Sinha and SC Basak, *J. Chem. Inf. Model.*, **46**, 1 (2006).

Mathematical descriptors of DNA sequences: Development and application, A Nandy, M Harle and SC Basak, *ARKIVOC*, **9**, 211-238 (2006).

On the dependence of a characterization of proteomics maps on the number of protein spots considered, M Randic, FA Witzmann, V Kodali and SC Basak, *J. Chem. Inf. Model.*, **46**, 116–122 (2006).

Optimal neighbor selection in molecular similarity: comparison of arbitrary versus tailored similarity spaces, BD Gute and SC Basak, *SAR QSAR Environ. Res.*, **17**, 37–51 (2006).

Predicting pharmacological and toxicological activity of heterocyclic compounds using QSAR and molecular modeling, SC Basak, D Mills, BD Gute and R Natarajan, in *Topics in Heterocyclic Chemistry, Vol. 3: QSAR and Molecular Modeling Studies of Heterocyclic Drugs*, SP Gupta, Ed., Springer-Verlag, Berlin, 39–80 (2006).

Prediction of tissue:air partition coefficients—theoretical versus experimental methods, SC Basak, D Mills and BD Gute, *SAR QSAR Environ. Res.*, **17**, 515–532 (2006).

Proteomics maps—Toxicity relationship of halocarbons studied with similarity index and genetic algorithm, M Vracko, SC Basak, K Geiss and F Witzmann, *J. Chem. Inf. Model.*, **46**, 130–136 (2006).

Quantitative structure-activity relationship modeling of juvenile hormone mimetic compounds for *Culex Pipiens* larvae, with a discussion of descriptor thinning methods, SC

Basak, R Natarajan, D Mills, DM Hawkins and JJ Kraker, *J. Chem. Inf. Model.*, **46**, 65–77 (2006).

Quantitative structure-toxicity relationships using chemodescriptors and biodescriptors, SC Basak, D Mills and BD Gute, in *Biological Concepts and Techniques in Toxicology: An Integrated Approach*, J Riviere, Ed., Marcel-Dekker, Inc., Taylor & Francis, New York, 61–82 (2006).

Similarity methods in analog selection, property estimation and clustering of diverse chemicals, SC Basak, BD Gute and D Mills, *ARKIVOC*, **9**, 157–210 (2006).

Statistical theory of spectra: Statistical moments as descriptors in the theory of molecular similarity, D Bielinska-Waz, P Waz and SC Basak, *Eur. Phys. J. B.*, **50**, 333–338 (2006).

Statistical theory of spectra as a tool in molecular similarity, D Bielinska-Waz, P Waz, SC Basak and R Natarajan, in *Symmetry, Spectroscopy, and SCHUR*, RC King, Ed., Nicolaus Copernicus University Press, Torun, 27–32 (2006).

2007

Graphical representation and numerical characterization of H5N1 avian flu neuraminidase gene sequence, A Nandy, SC Basak and BD Gute, *J. Chem. Inf. Model.* **47**, 945–951 (2007).

Information-theoretic biodescriptors for proteomics maps: Application to rodent hepatotoxicity, SC Basak, BD Gute, KT Geiss and FA Witzmann, in *Computation in Modern Science and Engineering, Proceedings of the International Conference on Computational Methods in Science and Engineering*, Volume 2, Part A, TE Simos and G Maroulis, Eds., American Institute of Physics, pp. 10–13.

A novel approach for the numerical characterization of molecular chirality, R Natarajan, SC Basak and TS Neumann, *J. Chem. Inf. Model.* **47**, 771–775 (2007).

Proper statistical modeling and validation in QSAR: A case study in the prediction of rat fat-air partitioning, SC Basak, D Mills, DM Hawkins and JJ Kraker, in *Computation in Modern Science and Engineering, Proceedings of the International Conference on Computational Methods in Science and Engineering 2007 (ICCMSE 2007)*, TE Simos, G Maroulis, Eds., American Institute of Physics, pp. 548–551.

Quantitative comparison of five molecular structure spaces in selecting analogs of chemicals, SC Basak, BD Gute and GD Grunwald, in *Computation in Modern Science and Engineering, Proceedings of the International Conference on Computational Methods in Science and Engineering*, Volume 2, Part A, TE Simos and G Maroulis, Eds., American Institute of Physics, pp. 544–547.

Quantitative structure-activity relationship (QSAR) modeling of juvenile hormone activity: Comparison of validation procedures, JJ Kraker, DM Hawkins, SC Basak, R Natarajan and D Mills, *Chemometr. Intell. Lab. Syst.* **87**, 33–42 (2007).

Quantitative structure-activity relationship (QSAR) studies of quinolone antibacterials against *M. fortuitum* and *M. smegmatis* using theoretical molecular descriptors, MC Bagchi, D Mills and SC Basak, *J. Mol. Modeling* **13**, 111–120 (2007).

A quantitative structure-activity relationship (QSAR) study of dermal absorption using theoretical molecular descriptors, SC Basak, D Mills and MM Mumtaz, *SAR QSAR Environ. Res.* **18**, 45–55 (2007).

Similarity studies using statistical and genetical methods, D Bielinska-Waz, P Waz and SC Basak, *J. Math. Chem.* **42**, 1003–1013 (2007).

Three dimensional structure-activity relationships (3D-QSAR) for insect repellency of diastereoisomeric compounds: A hierarchical molecular overlay approach, SC Basak, R Natarajan, W Nowak, P Miszta and JA Klun, *SAR QSAR Environ. Res.* **18**, 237-250 (2007).

2008

Mathematical biodescriptors of proteomics maps: Background and significance, SC Basak and BD Gute, *Curr. Opin. Drug Disc. Dev.* **11**, 320-326 (2008).

In press

Predicting bioactivity and toxicity of chemicals from mathematical descriptors: A chemical-cum-biochemical approach, SC Basak, D Mills and BD Gute, in *Advances in Quantum Chemistry*, DJ Klein and E Brandas, Eds., Elsevier–Academic Press.

Quantitative structure-activity relationship (QSAR) modeling of human blood:air partitioning with proper statistical methods and validation, SC Basak, D Mills, DM Hawkins and JJ Kraker, *Chemistry & Biodiversity*.

Use of mathematical structural invariants in analyzing combinatorial libraries: A case study with Psoralen derivatives, SC Basak, D Mills, BD Gute, AT Balaban, K Basak, GD Grunwald, in *Some Aspects of Mathematical Chemistry*, DK Sinha, SC Basak, RK Mohanty and IN Basumallick, Eds., Visva–Bharati University, India.

Use of proteomics based biodescriptors in the characterization of chemical toxicity, Z Bajzer, SC Basak, M Vracko Grobelsek and M Randic, in *Genomic and Proteomic Applications of Toxicity Testing*, MJ Cunningham, Ed., Humana Press, Inc.: Totowa, NJ.

Variable molecular descriptors, M Randic and SC Basak, in *Some Aspects of Mathematical Chemistry*, DK Sinha, SC Basak, RK Mohanty and IN Basumallick, Eds., Visva–Bharati University, India.

Submitted

Characterization of toxicoproteomics maps for mixtures and individual toxicants using information theoretic approach, SC Basak, BD Gute, N Monteiro-Riviere and FA Witzmann, in *Principles and Practice of Mixture Toxicity*, MM Mumtaz, Ed., Wiley-VCH Weinheim.

Mathematical chemistry and chemobioinformatics: A holistic view involving optimism, intractability, and pragmatism., SC Basak and D Mills, in *Proceedings of the 22nd International*

Course & Conference on the Interfaces among Mathematics, Chemistry & Computer Sciences, A Graovac and I Gutman, Eds.

Molecular overlay as a tool to model bio-specificity: A case study with mosquito repellents, R Natarajan and SC Basak, in *Lecture Notes of the First Indo-US Lecture Series on Discrete Mathematical Chemistry*, SC Basak and R Balakrishnan, Eds.

Molecular similarity: Defining, quantifying and tailoring structure spaces, BD Gute and SC Basak, in *Lecture Notes of the First Indo-US Lecture Series on Discrete Mathematical Chemistry*, SC Basak and R Balakrishnan, Eds.

NMR spectral invariants — A new class of descriptors for diastereomers and enantiomers, R Natarajan and SC Basak, *Croat. Chem. Acta.*

Predicting chemical reactivity and bioactivity from structure: A mathematical-cum-computational approach, SC Basak, D Mills, R Natarajan and BD Gute, in *Theory of Chemical Reactivity*, PK Chattaraj, Ed., Taylor & Francis.

Quantitative structure-activity relationship (QSAR) modeling of human blood:air partitioning with proper statistical methods and validation, SC Basak, D Mills, DM Hawkins and JJ Kraker, *Drug Metab. Lett.*

Quantitative structure-activity relationship modeling of mosquito repellents using calculated descriptors, R Natarajan, SC Basak, D Mills, JJ Kraker and DM Hawkins, *Croat. Chem. Acta.*

Use of graph invariants in the protection of human and ecological health, SC Basak, D Mills and MM Mumtaz, in *Lecture Notes of the First Indo-US Lecture Series on Discrete Mathematical Chemistry*, SC Basak and R Balakrishnan, Eds.

VII. Interactions/Transitions

a) Participation at Meetings

1. Dr. Basak delivered the invited lecture *Predicting bioactivity and toxicity of chemicals from computational chemistry and mathematical proteomics* at the Conferentia Chemometrica 2005 in Hajdúszoboszló, Hungary, August 28–30.
2. Basak gave two invited lectures at the International Conference of Computational Methods in Sciences and Engineering 2005 in Korinthos, Greece, October 21-26: *Use of proteomics-based biodescriptors versus chemodescriptors in predicting halocarbon toxicity: An integrated approach* and *A comparative study of arbitrary versus tailored molecular similarity metrics in property/toxicity/bioactivity prediction* both authored jointly by Basak, Brian Gute (NRRI) and Douglas M. Hawkins (School of Statistics, University of Minnesota Twin Cities).
3. Basak and colleagues attended the Computational Methods in Toxicology and Pharmacology conference held in Shanghai, October 29 – November 1, 2005. The

group presented the following papers:

i. Basak presented the paper *The role of chemodescriptors and biodescriptors in predicting bioactivity and toxicity*, authored jointly with Gute and Hawkins (U of MN Twin Cities).

ii. Gute presented the collaborative research paper *Property specific tailoring of molecular similarity metrics*, authored jointly with Hawkins and Basak.

4. Basak and Natarajan gave the following presentations at an international conference on Drug Discovery Based on Darjeeling Area Biodiversity, held November 7, 2005:

   i. *Combining modern drug discovery methods with biodiversity of Darjeeling plants in the discovery of pharmaceuticals and cosmeceuticals*, by Basak.

   ii. *Biodiversity of western ghats*, by Natarajan.

5. Basak and Natarajan gave the following presentations in a mini-symposium at a conference on Current Advances in QSAR Studies in Kolkata, India, held November 8, 2005:

   i. *QSAR Modeling: Descriptor thinning and cross validation*, authored by Ramanathan Natarajan and Basak.

   ii. *Advancing frontiers of mathematical chemistry*, by Basak.

6. Gute presented the paper *Use of proteomics-based mathematical biodescriptors in characterizing chemical toxicity*, authored jointly with Basak, at the 2005 Scientific Conference on Chemical and Biological Defense Research in Baltimore, MD, November 14- 16, 2005.

7. While attending the 2005 Scientific Conference on Chemical and Biological Defense Research, Gute discussed potential joint toxicoproteomics research with collaborators from Vital Probes, Inc., a company focused on developing technology for the detection of and defense against biological agents and infectious diseases.

8. Basak gave an invited lecture entitled *The utility of mathematical descriptors in the prediction of property, biochemical activity and toxicity of chemicals* at the Department of Biochemistry, University of Calcutta, Kolkata, India, November 21, 2005.

9. Basak gave an invited presentation entitled *Theoretical descriptor-based QSARs in predicting skin penetration of chemicals*, authored with Jim Riviere (North Carolina State University, Raleigh), Ronald Baynes (North Carolina State University, Raleigh), and Gute at the AFOSR JP-8 Jet Fuel Toxicology Meeting, University of Arizona, Tucson, November 30–December 2, 2005.

10. Basak and Gute visited North Carolina State University, Raleigh, to participate in the Jet Fuel Meeting "Integrating models on lung disposition and toxicokinetics of jet fuels" organized at the Mechanical and Aerospace Engineering Department of NCSU. Basak gave an invited presentation entitled *Clustering and QSAR of JP8 fuel chemicals*.

11. Basak and collaborators presented the following papers at the 12th International Workshop on Quantitative Structure-Activity Relationships in Environmental Chemistry (QSAR 2006), May 8–12, in Lyon, France:

    i. *Prediction of halocarbon toxicity using chemodescriptors and proteomics based biodescriptors: An integrated approach,* by Basak.

    ii. *Prediction of dermal absorption using quantitative structure-activity relationship modeling,* authored collectively by Basak, Denise Mills (NRRI) and Moiz Mumtaz and Selene Chou (both of the Agency for Toxic Substances and Disease Registry).

    iii. *Use of tailored similarity in estimating toxicity of chemicals,* authored collectively by Gute, Basak and Douglas Hawkins (University of Minnesota Twin Cities).

    iv. *A tailored approach to clustering and data mining,* authored collectively by Gute, Basak and James Riviere (North Carolina State University).

    v. *Map information content: An information-theoretic biodescriptor for characterizing toxic response in proteomics maps,* authored collectively by Gute, Basak and Frank Witzmann (Indiana University School of Medicine).

    vi. *Predicting toxicity of uncouplers of oxidative phosphorylation: A hierarchical QSAR approach,* authored jointly by Natarajan, Basak, Megan Forbes (NRRI), and Jessica Kraker and Douglas Hawkins (both from University of Minnesota Twin Cities).

    vii. *Developing QSAR models using primary and/or secondary descriptors,* authored jointly by Natarajan, Mills, Basak, Kraker and Hawkins.

    viii. *NMR spectral invariant: Novel descriptors for diastereomers and enantiomers,* authored jointly by Natarajan and Basak.

    ix. *Proper use of cross-validation while descriptor thinning: Naïve versus true q-square,* authored collectively by Natarajan, Basak, Kraker and Hawkins.

    x. *Mathematical bio-descriptors of DNA sequence structure and their applications,* authored jointly by Ashesh Nandy (NRRI/visiting scientist) and Basak.

    xi. *Graphical representation and numerical characterization of H5N1 avian flu neuraminidase gene sequence,* authored jointly by Nandy, Basak and Gute.

    xii. *QSAR checking and validation,* authored collectively by Hawkins, Basak, Kraker, and Mills.

    xiii. *Biophoric overlay of diastereomers using Hartree Fock and density functional theory optimized structures. 3D-QSAR to predict insect repellency,* authored collectively by Basak, Natarajan, Przemyslaw Miszta (N. Copernicus University, Poland) and Jerome Klun (USDA Agricultural Research Service).

    xiv. *A novel approach for the numerical characterization of molecular chirality,* authored collectively by Terrence Neumann (UMD Chemistry Department graduate student), Natarajan and Basak.

12. Basak and Natarajan presented the following papers at the meeting of the International Academy of Mathematical Chemistry (IAMC), June 15–17, Dubrovnik,

Croatia:

i. *Mathematical descriptors in the classification/ prediction of biological processes,* an invited talk presented by Basak.

ii. *On hierarchical QSAR approach,* co-authored by Basak and presented by Natarajan.

13. Basak and Natarajan presented the following papers at the 21st Dubrovnik International Course and Conference on the interfaces among Mathematics, Chemistry and Computer Sciences (Math/Chem/Comp 2006), June 19–24, Dubrovnik, Croatia:

i. *Prediction of tissue partition co-efficients using mathematical structural descriptors versus experimental properties,* authored jointly by Basak, Denise Mills, and Brian Gute (all from NRRI).

ii. *Descriptor thinning and proper cross-validation in QSAR,* authored jointly by Natarajan, Basak, and Jessica J. Kraker and Douglas M. Hawkins (both from the Department of Applied Statistics, U of MN).

14. Dr. Basak and Ramanathan Natarjan gave the following invited lectures on October 27, 2006 at a seminar on mosquito control entitled A Novel Approach to Designing Bioactive Compounds and Repellents:

i. *Computer assisted design of novel insect repellents,* authored jointly by Natajaran and Basak.

ii. *Mathematical descriptor based approaches to chemical design,* authored by Basak.

15. Basak gave an invited lecture entitled *Applications of mathematical chemistry in modern drug discovery and environmental protection* at the Ramakrishna Mission Vivekananda University, Kolkata, India, on November 2, 2006.

16. Basak and collaborators gave the following presentations at the First Indo-US Lecture Series on Discrete Mathematical Chemistry, Bangalore, Tamil Nadu, India, January 8–11, 2007:

i. *Advancing frontiers of mathematical chemistry,* by Basak.

ii. *Statistical tools for building robust QSAR models,* by Jessica Kraker (University of Wisconsin-Eau Claire).

iii. *Hierarchical biophore overlay as a 3-D QSAR approach,* by Natarajan.

iv. *Molecular similarity: Defining, quantifying, and tailoring structure spaces,* by Brian D. Gute.

17. Basak, Jim Riviere and Ronald Baynes (both of North Carolina State University, Raleigh), Gute, and Frank A. Witzmann (Indiana University School of Medicine) presented the paper *Predicting skin penetration and interaction with JP-8* at the JP-8 Jet Fuel Toxicology conference organized by the U.S. Air Force, in Tucson, AZ, January 17–19, 2007.

18. Basak gave an invited keynote lecture entitled *Chemo-bioinformatics: Characterization of molecules and biomolecules using mathematical descriptors* at the 15th International Symposium on Spectroscopy in Theory and Practice, April 18–21, 2007, Nova Gorica,

Slovenia.

19. Basak presented two invited lectures at the National Institute of Chemistry, Ljubljana:
    i. *DNA and proteomics-based biodescriptors and their applications*, on April 23, 2007.
    ii. *Development and applications of chemodescriptors*, on April 24, 2007.
20. Basak presented the lecture *Mathematical descriptors of proteomics maps and their biological applications* at the meeting of the International Academy of Mathematical Chemistry (IAMC), June 7–10, 2007, Dubrovnik, Croatia.
21. Basak attended the 22nd International Course & Conference on the Interfaces Among Mathematics, Chemistry & Computer Sciences, June 11-16, 2007, Dubrovnik, Croatia, where he gave an invited lecture and a two-lecture short course:
    i. *Use of arbitrary and tailored molecular similarity methods in the estimation of property/ bioactivity of chemicals* (invited lecture), authored jointly by Basak, Gute, Natarajan and Denise Mills.
    ii. *Chemodescriptors and biodescriptors: Mathematical basis and applications* (short course), authored by Basak.
22. Basak and collaborators gave the following invited lectures at the Second Indo-US Lecture Series on Discrete Mathematical Chemistry, Kalpetta, Kerala, India, June 20–25, 2007:
    i. *Background and history of the Mathematical Chemistry Lecture Series*, inaugural lecture by Basak.
    ii. *Mathematical structure descriptors: Development and applications in chemistry, drug discovery, environmental protection, and bioinformatics*, authored by Basak.
    iii. *Molecular similarity methods in property prediction*, authored jointly by Gute and Basak.
    iv. *Numerical characterization of molecular chirality*, by Natarajan and Basak.
    v. *Realizing a balance via mathematical chemistry*, valedictory lecture by Basak.
    vi. *Towards comparative genomics: Numerical descriptors for DNA sequences*, by Ashesh Nandy (Jadavpur University, Jadavpur, India).
    vii. *Some basic approaches in computational chemistry*, by Gute.
23. Basak delivered the first A. N. Bhaduri Memorial Lecture entitled *Modeling in drug design and environmental protection* at the B.C. Guha Center for Genetic Engineering and Biotechnology, University of Calcutta, on July 3, 2007, Kolkata, West Bengal, India.
24. Basak and Denise Mills participated in the Computational Chemistry Maui Workshop 2007 organized by the Defense Threat Reduction Agency, Chemical and Biological Technologies and Threat Agency Sciences, US Department of Defense, 13–16 August, 2007. Bask gave two invited lectures at the workshop:
    i. *Estimation of tissue partitioning of chemicals from their structure: A hierarchical QSAR approach*, authored jointly by Basak and Mills.

ii. *Mathematical biodescriptors of proteomics maps and their toxicological applications,* authored jointly by Basak and Gute.

25. Basak and co-workers Gute, Douglas Hawkins, and Natarajan traveled to Boston, MA, where they gave the following presentations at the 234th ACS National Meeting, August 19–23, 2007:

i. *Predicting allergic contact dermatitis: alternative statistical approaches to chemical classification,* authored jointly by Basak, Mills and Hawkins (School of Statistics, TC campus, U of MN); invited lecture, Division of Computers in Chemistry

ii. *Use of theoretical descriptors in predicting aryl hydrocarbon (Ah) receptor binding affinity of dibenzofurans: a hierarchical QSAR approach,* authored jointly by Basak and Mills; Division of Chemical Toxicology

iii. *Prediction of blood-air and tissue-air partition coefficients: Calculated molecular descriptors versus experimentally determined properties,* authored jointly by Basak and Mills; Division of Computers in Chemistry

iv. *Molecular overlay as a tool to model bio-specificity: A case study with mosquito repellents,* authored jointly by Natarajan and Basak; Division of Agrochemicals

v. *Relative chirality index: Novel approach for the numerical characterization of molecular chirality,* authored jointly by Natarajan and Basak as an invited lecture; Division of Chemical Information

vi. *Proper use of cross-validation while descriptor-thinning: Naïve versus true* q^2, authored jointly by Basak, Natarajan, Hawkins and Kraker; Division of Chemical Information

vii. *Mutagen/non-mutagen classification of diverse and structurally homogenous chemicals using calculated molecular descriptors: a hierarchical approach,* authored jointly Basak, Mills and Hawkins; Division of Chemical Toxicology

viii. *Development of mathematical biodescriptors for proteomics maps,* authored jointly by Gute and Basak; Division of Chemical Information

ix. *Tailoring molecular similarity metrics for property estimation,* authored jointly by Gute and Basak; Division of Chemical Information

x. *Mathematical biodescriptors for DNA sequences: Applications to avian influenza,* authored jointly Nandy, Gute and Basak; Division of Biological Chemistry

xi. *QSAR approach to modeling membrane permeability,* authored jointly by Basak and Gute; Division of Computers in Chemistry

xii. *QSAR model assessment,* authored by Hawkins and Kraker; Division of Computers and Chemistry

26. Basak attended the Mathematical Methods in Chemistry 2007 (MCC 2007) conference, September 22-24, organized by the University of Split and the Rudjer Boskovic Institute, Croatia where he gave the following presentations:

i. "Mathematical descriptors of chemical and biological systems: Development and applications" authored jointly by Basak, Gute and Mills.

ii. "Hierarchy of knowledge creation via mathematical chemistry" authored by Basak.

27. Basak traveled to Corfu, Greece, to attend the Mathematical Chemistry Symposium of the International Conference on Computational Modeling in Science and Engineering, 2007 (ICCMSE 07) and give the following presentations:

i. "Proper statistical modeling and validation in QSAR: A case study in the prediction of rat fat; air partitioning" authored jointly by Basak, Mills, Hawkins and Kraker.

ii. "Quantitative comparison of five molecular structure spaces in selecting analogs of chemicals" authored jointly by Basak, Gute and Hawkins.

iii. "Information theoretic biodescriptors of proteomics maps" author jointly by Basak and Gute.

iv. "Similarity and dissimilarity of DNA/ RNA sequences" authored jointly by D. Bielinska-Waz (Instytut Fizyki, Uniwersytet Mikolaja Kopernika, Torun, Poland), P. Waz (Centrum Astronomii, Uniwersytet Mikolaja Kopernika, Torun, Poland), W. Nowak (Department of Physics, University of Torun), A. Nandy (School of Environmental Science, Jadavpur University, and Kolkata, India) and Basak.

28. Basak traveled to Tiruchirappalli, Tamil Nadu, India, to organize the Third Indo-US Lecture Series on Discrete Mathematical Chemistry (Special Lectures on Cheminformatics and Bioinformatics) January 7–10, 2008, as the US Chairperson. The binational, USA-India event was organized under the joint auspices of the Natural Resources Research Institute (NRRI) and Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli. The NRRI team gave the following lectures at the conference:

i. "Mathematical chemodescriptors and biodescriptors: Development and applications" by Basak.

ii. "An integrated chemo-bioinformatic approach to bioactivity/ toxicity prediction" by Basak.

iii. "Realizing a balance via mathematical chemistry" by Basak at the concluding valedictory session of the lecture series.

iv. "Mathematical characterization of chirality: The hallmark of life's chemistry" by Natarajan and Basak.

v. "Characterizing molecular similarity and similarity methods" by Gute and Basak.

29. Basak and Gute continued on to Kolkata, India, where Basak gave an invited lecture entitled "Mathematical structural invariants: Development and applications in predicting property/bioactivity/toxicity of chemicals" at the Department of Biophysics and Molecular Biology, University College of Science, University of Calcutta, Kolkata, India, on January 14, 2008. Basak and Gute also participated in a discussion session

with attendees of the lecture on various aspects of chemoinformatics, bioinformatics, and computational biology research in Eastern India.

30. Basak and coworkers organized the joint meeting of the 5th Indo-US Workshop on Mathematical Chemistry and the 8th International Conference on Mathematical Chemistry, held in Duluth, MN at the University of Minnesota Duluth, June 22–27, 2008. Basak and coworkers also gave a number of presentations at the meeting:
    i. "Chemo-bioinformatics: An integration of molecular structure and 'omics' based approaches for predicting bioactivity" by Basak.
    ii. "Numerical characterization molecular chirality" by Natarajan.
    iii. "Recent directions in predictor selection" coauthored by Kraker and Hawkins.
    iv. "Tailoring molecular similarity methods to optimize activity estimation" coauthored by Gute, Basak and Hawkins.
    v. "Prediction of biological partition coefficients: Calculated molecular descriptors vs experimentally determined properties" coauthored by Mills, Basak, Gute and Moiz M. Mumtaz (Agency for Toxic Substances and Disease Registry, Atlanta, GA).
    vi. "Computer-assisted design of chelating mineral collectors" by Natarajan.
    vii. "DNA sequence descriptors based on information theory" coauthored by Natarajan, Ramamurthy Jayalakshmi (Bharathidasan University, Tiruchirappalli, India), M Vivekanandhan (Bharathidasan University, Tiruchirappalli, India), Ganapathy Natarajan (University of Minnesota Duluth), and TM Anbazhagan (Crux fusion, Bangalore, India).
    viii. "Numerical characterization studies of mutations among neuraminidase gene sequences of the H5N1 avian flu strains from 1997 to 2007" coauthored by Gute, Ashesh Nandy (Jadavpur University, Kolkata, India), Ambarnil Ghosh (Jadavpur University, Kolkata, India) and Basak.

b) Consultative and Advisory Functions

1. Dr. Basak's research team has been consulting with Dr. Jim Riviere, another AFOSR grantee, and his colleagues Ronald Baynes and Xin-Rui Xia. Dr. Riviere's group is currently testing a selection of JP-8 constituents for skin penetration using a membrane coated fiber (MCF) method. Dr. Basak's team has consulted with them on the selection of test chemicals to better explore the JP-8 chemical structure space. An iterative process has been agreed upon, and the consultation will continue into the near future. Once results have been generated, the data will be used by Dr. Basak's team to develop models for dermal penetration. As part of this collaboration, Basak and Brian Gute visited Dr. Riviere's laboratory in January to discuss progress of the experimental work and future directions.
2. Basak discussed the potential for joint toxicoproteomics research with collaborators at the University of Calcutta and anti-tuberculosis drug design with collaborators at the Indian Institute of Chemical Biology, Kolkata, November 21, 2005.

3. Basak's research team continued its collaborative work with Dr. Jeff Fisher. Fisher's (AFRL) studies of JP-8 pharmacokinetics in animals will be used to develop computational models to predict the experimental data. Successful chemical structure based models will assist in the estimation of pharmacokinetics parameters for untested JP8 chemicals.

4. Basak and Ramanathan Natarajan traveled to London to discuss collaborative research with colleagues at the Department of Crystallography, Birkbeck College, and University of London.

5. Basak traveled to Charleston, SC, to discuss computational approaches to cancer research and toxicology and the potential for future collaborative research with colleagues at the Medical University of South Carolina.

6. Dr. Basak's research team has been consulting with BioPred, a small biological modeling company, in the field of prion research.

7. Basak and Gute have been consulting with researchers from VitalProbes, a small east coast bio-tech company. It is hoped that this interaction will lead to a direct transfer of biodescriptor technology to a corporate end-user.

8. Basak, the current President of the International Society of Mathematical Chemistry (ISMC), discussed matters related to the progress of ISMC with International Academy of Mathematical Chemistry members at the 2007 meeting of the International Academy.

9. Basak and Gute traveled to Amherst, MA, to attend the 6th and 7th International Hormesis Conferences organized by Dr. Edward Calabrese and to discuss continued collaborative research and proposal development on hormesis and anti-cancer drug design with Dr. Calabrese at the University of Massachusetts, Amherst.

10. Basak established ties with both the Indo-US Science and Technology Forum and the Department of Science and Technology (Gov't. of India), and continues to collaborate with both groups to organize and provide funding for programs to benefit young scientists in India. These collaborations have resulted in the creation of the Indo-US Lecture Series on Discrete Mathematical Chemistry. Two of these lecture series were held in India in 2007 (Bangalore and Kalpeta), a third was held in January of 2008 (Tiruchirappalli, India), and a fourth is being planned for January 2009 (Hyderabad, India).

11. Basak interacted with groups on the University of Minnesota Duluth campus regarding the creation of an undergraduate program in chemo- and bioinformatics.

12. Basak discussed the organization of a new Chemo-bioinformatics and Computational Biology Department with the Vice Chancellor of the Ramakrishna Mission Vivekananda University in Kolkata, India.

c) Transitions

BioPred Consulting

Customer:
BioPred (Computational Bioactivity Prediction Company, LLC)
2150 Analysis Drive
Bozeman, Montana 59718

Contact: Timothy Nagel, Company President
Tel: (406) 582-0005

Result: POLLY, the molecular descriptor calculation software (developed by Basak et al and copyright of the University of Minnesota), has been augmented to calculate descriptors for organometallic compounds. Organometallics are known to be active in inhibiting the formation of prion proteins, misfolded proteins involved in diseases such as Mad Cow disease and scrapies, and the ability to predictively model these compounds would be useful in developing new or improved treatments for these diseases.

In addition to the revisions made to the *POLLY* software last year for BioPred, Dr. Basak and colleagues have created and provided to the scientists at BioPred a virtual library of potential anti-prion compounds based upon lead structures identified by the BioPred scientists. Over 98,000 virtual structures were created and molecular indices calculated by *POLLY v2.3* were provided to BioPred scientists for further modeling and lead identification.

Application: A company, BioPred in Bozemann, Montana, has used POLLY in the design of chemicals active against prions, misfolded proteins involved in diseases such as Mad Cow disease. Their preliminary results are show promise in developing new treatments for Mad Cow disease. In addition, a system capable of calculations on organometallics could be useful in developing new ligands for biomedicinal applications or environmental remediation processes.

VIII. New Discoveries, Inventions, or Patent Disclosures

This project resulted in some exciting new developments, spurring the development of two novel fields of research that are rapidly growing and gaining acceptance within the scientific community, and has led to new discoveries fundamental to the continued development of the field of quantitative structure-activity relationship modeling.

The primary focus of this project, the development of mathematical invariants (biodescriptors) for proteomic maps, has been key to the development of the field of mathematical proteomics. Research initially published under an earlier AFOSR project in 2001 formed the foundation for the rapidly growing field of mathematical proteomics and the application of graph theory and information theory to two-dimensional, biological data. While

the work was initiated specifically to characterize proteomics maps, the fundamental nature of this work makes it useful for characterizing any type of two-dimensional biological data—especially that resulting from any kind of physical separation technique. However, the work pioneered by Basak's group in mathematical proteomics is not limited solely to applications related to two-dimensional gels. These descriptors and invariants can be applied to other types of biological data, including proteomics data from MALDI and SELDI type analyses. With the large amount of data generated by these continually evolving and changing modern techniques, methods to characterize that data into a smaller set of invariants will be extremely useful in the future.

Likewise, Basak's research on invariants to characterize DNA and RNA sequence data has led to the development of another novel field of research, using mathematical techniques to characterize complex biological sequences. This emerging field has caught on quickly, and scientists around the world are working to develop useful methods for characterizing DNA primary sequence data. As shown in the recent paper by Nandy, Basak and Gute, these techniques will provide tools to distinguish between various strains of viruses and bacteria, helping scientists to rapidly determine the virulence of a mutant strain before it can develop into an epidemic.

Finally, Dr. Basak's research team has pursued the concept of integrated quantitative structure activity relationship (I-QSAR) studies. This approach proposes that combining structural information about a chemical with biological response data should improve the ability to predict a biologically-relevant endpoint. At this stage, preliminary results from statistical modeling done both in Dr. Basak's laboratory and by Dr. Hawkins show that this is indeed the case—integrated models using both chemodescriptors and biodescriptors demonstrate an improved capacity to accurately model a biological response.

IX. Honors and Awards

a) Honors

1. Dr. Basak co-authored a paper cited as "Most Accessed" by the *Journal of Chemical Information and Modeling* for January to March, 2006. The article, "Combining chemodescriptors and biodescriptors in quantitative structure-activity relationship modeling," **46**, 9–16 (2006), was authored jointly by Douglas M. Hawkins (University of Minnesota Twin Cities), Basak, Jessica Kraker (University of Minnesota Twin Cities), Kevin Geiss (Air Force Research Laboratory) and Frank A. Witzmann (Indiana University School of Medicine).

2. Dr. Basak was named one of the authors most cited by scientists in India for research in 2003–05. The study, "Assessment of India's Research Literature" was conducted for the Defense Technical Information Center in Fort Belvoir, Virginia, by Ronald N. Kostoff, Dustin Johnson, Christine Bowles, and Simha Dodbele.

3. Dr. Basak was the guest editor of the special issue of the American Chemical Society's *Journal of Chemical Information and Modeling* which published papers

presented at the Fourth Indo-U.S. Workshop on Mathematical Chemistry with Applications to Drug Design, Risk Assessment of Chemicals, Cheminformatics, Bioinformatics, Computational Biology and Toxicology.

4. Dr. Basak was invited to join the editorial advisory board of the international journal *Current Computer-Aided Drug Design* (Bentham Science Publishers).
5. Basak was also invited to join the editorial board of the international journal *Open Medicinal Chemistry* (Bentham Science Publishers).
6. Basak was invited to be a keynote speaker at the 15th International Symposium on Spectroscopy in Theory and Practice, April 18–21, 2007, organized at Nova Gorica, Slovenia, by the Slovenian Chemical Society, in collaboration with the Jožef Stefan Institute, National Institute of Chemistry, Slovenia.
7. Basak delivered the first A.N. Bhaduri Memorial Lecture at the B.C. Guha Center for Genetic Engineering and Biotechnology, University of Calcutta, organized jointly by Science for Society (India) in collaboration with the Eastern India Chapter of the International Society of Mathematical Chemistry on July 3, 2007, Kolkata, West Bengal, India.

b) Advisory/Organizational Positions Held by Dr. Basak

1. President of the International Society for Mathematical Chemistry (2003–2007).
2. Editorial board member of the international journal *SAR and QSAR in Environmental Research* (Gordon and Breach).
3. Editorial board member of the international journal *Current Computer-Aided Drug Design* (Bentham Science Publishers).
4. Editorial board member of the international journal *Open Medicinal Chemistry* (Bentham Science Publishers).
5. Dr. Basak chaired a session at Conferentia Chemometrica 2005 in Hajdúszoboszló, Hungary, August 28–30.
6. Basak chaired a session on Mathematical Chemistry and QSAR at the International Conference of Computational Methods in Sciences and Engineering 2005" in Korinthos, Greece, October 21–26.
7. Basak co-chaired a QSAR and Predictive Molecular Modeling session at the Computational Methods in Toxicology and Pharmacology conference held in Shanghai, October 29–November 1, 2005.
8. Basak organized an international conference on Drug Discovery Based on Darjeeling Area Biodiversity, held November 7, 2005.
9. Basak organized a mini-symposium at a conference on Current Advances in QSAR Studies in Kolkata, India, held November 8, 2005.

10. Basak chaired a session on Environmental Fate Modeling at the 12th International Workshop on Quantitative Structure-Activity Relationships in Environmental Chemistry (QSAR 2006) in Lyon, France, May 8–12.

11. Basak chaired the scientific session "Modeling of Bioactive Molecules" at the 21st Dubrovnik International Course and Conference on the interfaces among Mathematics, Chemistry and Computer Sciences (Math/Chem/Comp 2006), Dubrovnik, Croatia, June 19–24.

12. Dr. Basak and co-workers organized the First Indo-US Lecture Series on Discrete Mathematical Chemistry held at the PES Institute of Technology in Bangalore, India, from January 8–11, 2007. Basak was one of the co-chairs for the event.

13. Basak chaired a session at the International Academy of Mathematical Chemistry (IAMC) 2007 conference held in Dubrovnik, Croatia, June 7–10, 2007.

14. Basak and co-workers organized the Second Indo-US Lecture Series on Discrete Mathematical Chemistry held at the Woodlands Hotel in Kalpetta, Kerala, India, from June 20–25, 2007. Basak was one of the co-chairs for the event.

15. Basak is a member of the International Scientific Advisory Board for the Fourth International Symposium on Computational Methods in Toxicology and Pharmacology Integrating Internet Resources, Moscow, Russia, September 1–5, 2007.

16. Basak and co-workers organized the Third Indo-US Lecture Series on Discrete Mathematical Chemistry held at the Hotel Royal Southern in Tiruchirappalli, Tamil Nadu, India, from January 7–10, 2008. Basak was one of the co-chairs for the event.

17. Basak and co-workers organized the joint meeting of the Fifth Indo-US Workshop on Mathematical Chemistry and Eighth International Conference on Mathematical Chemistry held at the University of Minnesota Duluth in Duluth, Minnesota, from June 22–27, 2008. Basak was one of the co-chairs for the event.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
07-08-2008	Final Performance Report	Jul 2005 -Jul 2008; Aug 2008

4. TITLE AND SUBTITLE
Predicting Chemical Toxicity from Proteomics and Computational Chemistry

5a. CONTRACT NUMBER

5b. GRANT NUMBER
G AF/FA9550-05-1-0456

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Subhash C. Basak, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Minnesota Duluth
Natural Resources Research Institute
5013 Miller Trunk Hwy.
Duluth, MN 55811

8. PERFORMING ORGANIZATION REPORT NUMBER
NRRI/TR-2008/18

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office of Scientific Research
4015 Wilson Boulevard
Room 413
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-OSR-VA-TR-2012-0489

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project focuses on a two-pronged approach to modeling toxicity data. A standard structure-based QSAR approach using chemodescriptors (descriptors based on the chemical structure of the toxicant) has been coupled with the development of biodescriptors, a novel set of mathematical descriptors derived from 2-DE proteomic gel analyses.

In the realm of chemodescriptors, the team has used them in a hierarchical manner, from topological and geometrical, to high-level quantum chemical indices such as vertical electron affinity which is believed to be mechanistically related to chemical toxicity.

Biodescriptor development has focused on three main approaches: global descriptors that characterize the entire proteomics map, local descriptors that characterize a subset of the proteins present in the gel, and spectrum-like descriptors. These efforts have been further enhanced through the use of robust statistical approaches and the development of new statistical techniques for analyzing the full set of proteins present in a proteomics map.

15. SUBJECT TERMS
Topological indices; chemodescriptors; proteomics; biodescriptors; 2-D gel electrophoresis; QSAR; hierarchical QSAR; integrated QSAR

16. SECURITY CLASSIFICATION OF: UNCLASSIFIED			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	UNLIMITED	23	Subhash C. Basak, PhD
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED			19b. TELEPHONE NUMBER *(include area code)* 218-720-4230