Int. J. Mol. Sci. 2014, 15, 1-x manuscripts; doi:10.3390/ijms150x000x

**OPEN ACCESS**

International Journal of  
Molecular Sciences

ISSN 1422-0067  
www.mdpi.com/journal/ijms

*Type of the Paper (Article, Review, Communication, etc.)*

**Title of the Paper (M\_Title)**

**Firstname Lastname 1,\*, Firstname Lastname 2,†, Firstname Lastname 3,† and   
Firstname Lastname 4**

1 Full Affiliation, Address; E-Mail: author2@email

2 Full Affiliation, Address; E-Mails: author3@email (F.L.); author4@email (F.L.);

**†**These authors contributed equally to this work.

**\*** Author to whom correspondence should be addressed; E-Mail: author1@email;   
Tel.: +1-111-111-111 (ext. 123); Fax: +1-111-111-112.

*Received: / Accepted: / Published:*

**Abstract:** This is the abstract section. One paragraph only (Maximum 200 words).

**Keywords:** keyword; keyword; keyword (3–10 keywords separated by semi colons)

Introduction:

Risk assessment of chemicals is often carried out using very little of no experimental data. In 1984, the National Research Council evaluated the availability of toxicity data on industrial chemicals and concluded that that many of these chemicals have very few or no test data [l]. Such a situation is prevalent also with substances identified in industrial effluent, hazardous waste sites and environmental monitoring surveys [2]. In USA, the Toxic Substances Control Act (TSCA) inventory currently has about 86, 000 entries and the list is growing by nearly 3000 per year [3, 4]. Of some 3000 chemicals submitted yearly to the United States Environmental Protection Agency (USEPA) for the premanufacture notification (PMN) review process, more than 50% have no experimental data at all, less than 15% have empirical mutagenicity data, and about 6% have experimental ecotoxicological and environmental fate data.

Mutagenicity is one important toxicity data essential both environmental protection as well as new drug discovery. Identification of potential mutagenicity of environmental pollutants is important for the protection of human and ecological health. In the realm of drug discovery, early detection of mutagenicity of drug candidates can help in the effective allocation of resources for the expensive drug discovery protocol which costs US $400 million to 2 billion per drug [5, 6]]. Experimental determination of mutagenicity of all possible candidate chemicals, both for environmental protection and drug design, can be very expensive. Therefore, potential mutagenicity of chemicals from quantitative structure-activity relationship (QSAR) models has been accepted for the evaluation of chemicals in lieu of experimental mutagenicity data [7].

QSARs for any biological endpoint of a set of chemicals can be formulated either from other experimental properties or molecular descriptors calculated from molecular structure without the input of any other experimental data. The frequently used descriptors for QSAR are topological indices, substructures, and quantum chemical descriptors [7, 8 ]. For large sets of structures, high level quantum chemical descriptors can be very resource intensive. Alternately, descriptors derived from topological aspects of chemical structures, e.g.; topological indices and different types of substructures, have found successful applications in numerous good quality QSAR studies. For a recent summary of the topic, please see the review by Basak [9].

There are a few critical needs for the development of QSARs: a) Good quality and sufficiently large property/activity/ toxicity database, b) Computed molecular descriptors which quantify aspects of molecular structure associated with the toxicity or biological activity of interest, and c) Proper methods for model building. In previous studies, Basak et al [8] reviewed the results of various QSAR studies where topostructural (TS), topochemical (TC), geometrical (3-D), and quantum chemical (QC) indices were used in QSAR building in a graduated manner. Results of such hierarchical QSARs (HiQSARs) showed that in most cases a combination of TS and TC indices gave the best models. The addition of 3-D or QC descriptors to the set of independent variables did not make much improvement in model quality. The currently available software like PaDEL [10], Dragon [11], MolconnZ [12], POLLY [13], APProbe [14] are capable of calculating a large number of topological and substructural chemodescriptors. But, in many cases the situation becomes “rank deficient” because there are many more descriptors as compared to the number of data points to be modelled. So, proper statistical methods of variable selection, QSAR formulation, and model validation have to be followed for best results.

This paper has a two-fold objective: 1) Use a relatively new method, called interactive two way clustering (ITC) in developing QSAR for the prediction of mutagenicity of two data sets, viz., a homogeneous set of 95 aromatic and heteroaromatic amines and a structurally diverse set of 508 chemicals, and 2) Use the HiQSAR approach formulated by Basak et al [8] in evaluating the relative effectiveness of TS, TC, 3-D, and QC descriptors in the development of high quality QSARs for these two data sets.

Main text paragraph.

Main text paragraph.

**2. Results and Discussion**

Main text paragraph.

Main text paragraph.

|  |  |
| --- | --- |
| [add an equation here; use MS Word or MathType equation function] | (1) |

(Note: all equations should be completed within a two column table with one line, centered, no boarders, as example see above).

Main text paragraph.

Main text paragraph.

*2.1. This is Subsection Heading*

Main text paragraph.

Main text paragraph.

2.1.1. This is Subsection Heading

Main text paragraph.

Main text paragraph.

**Table 1.** Add a descriptive label of the table here.

[add the table here; use MS Word’s table function]

**Figure 1.** (**a**) Add a descriptive label of the figure here. (**b**) Add a descriptive label of the figure here. (**c**) Add a descriptive label of the figure here.

[add the figure here]

**3. Experimental Section**

3.1. Data

The dataset used in our analysis is due to Debnath *et al* [?] and concerns mutagenic activities of 95 aromatic and heteroaromatic amines in *S. typhimurium* TA98+S9 microsomal preparations. There are 275 descriptors for each compound: among them 97 topostructural, 162 topochemical, 10 3-dimensional and 6 quantum-chemical. Number of revertants per nmol of test culture in log scale (**log *R***) is the original response variable. For binary classification we take the 0/1 indicator **log *R*** being > 0 or < 0 as the response variable in our analysis.

The results obtained are compared with previous studies on a second, and more diverse, dataset records Ames mutagenicity of 508 chemical compounds (256 mutagens and 252 non-mutagens), and is taken from the CRC Handbook of Identified Carcinogens and Non-carcinogens [?]. The set of descriptors for these compounds includes the above type of descriptors, as well as a large number of Atom-pair (AP) descriptors.

3.2. Methods

3.2.1. Variable selection

3.2.2. Hierarchical QSAR and predictive models

**4. Conclusions**

Main text paragraph.

Main text paragraph.

**Acknowledgments**

Main text paragraph.

**Author Contributions**

Main text paragraph.

**Conflicts of Interest**

State any potential conflicts of interest here or “The authors declare no conflict of interest”.

**References and Notes**

[I ] National Research Council (1984) Toxicity Testing Strategies to Determine Needs and Priorities, National Academy Press, Washington, DC.

[2] Auer, CM., Nabholz, J.V. and Baetcke, K.P. (1990) Mode of action and the assessment of chemical hazards in the presence of limited data: use of structure-activity relationships (SAR) under TSCA, Section 5. Environ. Health Perspect. 87, 183-197.

[3] Toxic Substances Control Act (TSCA) (1976) Public Law 94-469, 90 Stat. 2003, 11 October 1976

[4] Cash, G., United States Environmental Protection Agency (USEPA, Personal communication, 2014.

[5] Adams, C.; Brantner, V. Estimating The Cost Of New Drug Development: Is It Really $802 Million? Health Aff. (Millwood), 2006, 25, 420-428.

[6] DiMasi, J.; Hansen, R.; Grabowski, H. The price of innovation: new estimates of drug development costs. J. Health Eco., 2003, 22, 151-185.

[7] M. Vracko, Mathematical (Structural) Descriptors in QSAR: Applications in Drug Design and Environmental Toxicology, In: Advances in Mathematical Chemistry and Applications, Subhash C. Basak, Guillermo Restrepo and J. L. Villaveces, Editors, Vol. 1,pp. 222-250, Bentham Science Publishers, in press, 2014.

[8] Basak, S. C., Mathematical Descriptors for the Prediction of Property, Bioactivity, and Toxicity of Chemicals from their Structure: A Chemical-Cum-Biochemical Approach, Current Computer-Aided Drug Design, 2013, 9, 449-462.

[9] Basak, S.C. Role of mathematical chemodescriptors and proteomics-based biodescriptors in drug discovery. Drug Dev. Res., 2010, 72, 1-9.

[10] PaDEL-Descriptor: <http://padel.nus.edu.sg/software/padeldescriptor/>

[11] DRAGON - Software for the Calculation of Molecular Descriptors, Version 5.4, 2006; Todeschini, R.; Consonni, V.; Mauri, A. et al.; Talete srl.; Milan, Italy

[12] MolconnZ, Version 4.05, 2003; Hall Ass. Consult.; Quincy, MA.

[13] Basak, S.C.; Harriss, D.K.; Magnuson, V.R. POLLY v. 2.3: 1988; Copyright of the University of Minnesota, USA.

[14] Basak, S.C.; Grunwald, G.D. APProbe. 1993; Copyright of the University of Minnesota, USA.

1. Author 1, A.B.; Author 2, C.D. Title of the cited article. *Journal Title* **2007**, *6*, 100–110.

2. Author 1, A.; Author 2, B. Title of the chapter. In *Book Title*, 2nd ed.; Editor 1, Editor 2, Eds.; Publisher: Publisher Location, Country, 2007; Volume 3, pp. 154–196.

3. Author 1, A.; Author 2, B. *Book Title*, 3rd ed.; Publisher: Publisher Location, Country, 2008;   
pp. 154–196.

4. Author 1, A.B.; Author 2, C. Title of Unpublished Work. Journal Abbreviation, phrase indicating stage of publication.

5. Author 1, A.B.; Author 2, C.D.; Author 3, E.F. Title of Presentation. In *Title of the Collected Work* (if available), Proceedings of the Name of the Conference, Location of Conference, Country, Date of Conference; Editor 1, Editor 2, Eds. (if available); Publisher: City, Country, Year (if available); Abstract Number (optional), Pagination (optional).

6. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.

7. Author 1, A.B.; Author 2, C.D. Title of the article. *Abbreviated Journal Name*, Year, Volume, (Page range), doi or other identification number. Available online: http://URL (accessed on Day Month Year).

8. Title of Site. Available online: http://URL (accessed on Day Month Year).

Reference list: We recommend the use of reference management software to prepare the references list (e.g., Endnote, http://www.mdpi.com/files/word-templates/MDPI.ens).

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).