Mathematical structural descriptors and mutagenicity assessment: A study with congeneric and diverse data sets

Subhabrata Majumdar, Subhash C. Basak, Gregory D. Grunwald, Mircea Diudea and Claudiu Lungu

Abstract: TBD later

Keywords: TBD later

1. INTRODUCTION

Hazard assessment of chemicals is often carried out in data poor situations [l]. The Toxic Substances Control Act (TSCA) Inventory, maintained by the United States Environmental Protection Agency (USEPA), currently has about 85,000 entries [2]. A large fraction of these chemicals has very little or no data needed for their hazard estimation [3]. The assessment of chemical mutagenicity is important both for environmental protection and drug discovery. Identification of potential mutagenicity for industrial chemicals and environmental pollutants is prerequisite to the protection of human and ecological health. For drug discovery, early mutagenicity detection for drug candidates can help in the effective allocation of resources in drug design protocol which costs on the average over US $2 billion [4].

Laboratory testing of mutagenicity for all possible candidate chemicals, can be very expensive. Therefore, assessment of potential mutagenicity of chemicals from Quantitative Structure-Activity Relationship (QSAR) models has been accepted for evaluation of chemicals in lieu of experimental mutagenicity data [5].

This paper has a two-fold objective: 1) Apply computed molecular descriptors in the formulation of QSARs for the prediction of mutagenicity of two data sets, viz., a homogeneous set of 95 aromatic and heteroaromatic amines and a large as well as structurally diverse set of 508 chemicals, and 2) Use a battery of various statistical and machine learning approaches in model building for mutagenicity assessment.

2. MATERIALS AND METHODS

2.1. Data (Subhabrata, please complete this section based on our earlier work) giving Table 1 to indicate diversity of the 508-chemical data set.

2.2. Descriptors

For this study we have used two collections of molecular descriptors. One set of descriptors, used frequently by the Cluj team of Diudea and collaborators, were calculated by the programs Schrodinger [6] and TopoCluj [7[8. More detailed references about these descriptors are given in Supplementary Tables 1 and 4. For the 95 and 508 data sets, ?? and ?? descriptors were calculated by Diudea et al.

The second set of molecular descriptors, used frequently by Basak et al, were calculated by the software POLLY [8], MolConnZ (9), Triplet [10], and MOPAC [11]. For the 95 and 508 chemical sets, ?? and ?? descriptors were calculated for this paper by this software.

2.3. Analysis of data by Statistical and machine learning methods

(Subhabrata, please complete this section)

3) RESULTS

(Subhabrata, please write the initial version and I will update that based on you PCA loading tables and others)

4) Discussion (To be completed later after results are written)

5) CONFLICT OF INTEREST

We confirm that there is no conflict of interest on the

content of this paper.

6) ACKNOWLEDGEMENTS

7) SUPPLEMENTARY MATERIAL

Supplementary material (Supplementary tables 1-4) is available on the publisher’s web site along with the published article.

8) REFERENCES:

1. National Research Council. Toxicity Testing Strategies to Determine Needs and Priorities, National Academy Press: Washington, DC, 1984.
2. Toxic Substances Control Act (TSCA) Inventory: <https://19january2017snapshot.epa.gov/tsca-inventory/about-tsca-chemical-substance-inventory_.html>; Accessed on April 11, 2018.
3. Auer, C.M.; Nabholz, J.V. and Baetcke, K.P. 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. 1990, 87, 183-197.
4. Innovation in the pharmaceutical industry: New estimates of R&DcostsJoseph A. DiMasia, Henry G. Grabowski, Ronald W. Hansen, (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. J. Health Econ. 47, 20–33
5. Quantitative structure-activity relationship (QSAR) Models for Mutagens and Carcinogens; Romaualdo Benigni, Ed.; CRC Press, Boca Raton, FL, 2003.
6. Schrodinger ref (Claudiu to provide)
7. TopCluj ref (Claudiu to provide)
8. S. C. Basak, D. K. Harriss and V. R. Magnuson, "POLLY v2.3," Copyright of the University of Minnesota, 1988.
9. MolconnZ v4.05, Quincy, MA: Hall Ass. Consult., 2003.
10. S. C. Basak, G. Grunwald and A. Balaban, "TRIPLET," Copyright of the Regents of the University of Minnesota, 1993.
11. J. Stewart, MOPAC Version 6.00, QCPE #455, Frank J. Seiler Research Laboratory: US Air Force Academy, CO, 1990.