QSAR modeling of toxicity of chemical mixtures

Abstract

The continued population growth, rapid urbanization, and simultaneous chemical advancement greatly increase chemical pollution worldwide. The chemical industries are flourishing in different parts of the world (mostly in developing countries) to fulfil the growing demand of mankind, and these socio-economical and industrial growths are creating a serious threat to environmental safety and ground-water prevention. The chemicals are mostly discharged through different sewage systems such as hospitals, cattle fields, aquacultures, etc.; and thus, a majority of chemical pollutants i.e., organic chemicals, plant protection products, pharmaceuticals, industrial, and household chemicals have been found in soil and water resources in ng/L to mg/L ranges. The bioactive chemicals i.e., pharmaceuticals, pesticides, insecticides, acaricides etc. are of major concern due to their bioactive nature. Antibiotics inevitably impact the lower environmental organisms (mostly aquatic and soil species). They mostly target the microorganism, especially the bacteria in the environment. Moreover, the bacterial species come in contact with multiple chemicals simultaneously, which may further potentiate the toxic effects. Different other species i.e., aquatic species (fish, algae, etc.), soil species (earthworm, termites, fungi, etc.), insets (honey bees etc.), etc. are also affected. Besides the environmental damage, the presence of such mixtures of toxic chemicals can cause serious damage to human beings due to bioaccumulation and biomagnification.

The pollutants are never found alone in the environment. Multicomponent chemical mixtures are formed continuously with variable combinations, and the number of such combinations is increasing. In some previous research articles, deviations of toxicity from the standard were reported due to significant interactions (additive response, synergism, antagonism) among chemicals. Besides this, the degradation products of chemicals also play an essential role in mixture toxicity as reported by Wang and co-workers. Therefore, a mixture toxicity assessment is a significant and necessary task which should be carried out for a proper risk assessment of chemicals against different ecological species. In the past decades, contamination of the aquatic environment with various pharmaceuticals and pesticides has grabbed the attention of global scientific communities and it is becoming an important topic of discussion after finding their simultaneous presence in water bodies. Scientists are now working on the risk assessment of chemical mixtures; however, mixture risk assessment by physical experimentation is not a good idea due to the unending possibilities of combinations. Besides this, the physical experiments are associated with numerous complications including ethical issues. Animal models for toxicity testing are time-consuming and costly as well as unethical. In 1959, Russell and Burch introduced the “3R” concept (Replacement, Reduction, and Refinement), to reduce the use of animals in scientific procedures.

There are two largely accepted and widely used mixture toxicity models, namely concentration addition (CA) and independent action (IA) which assume the additive nature of mixture components, but they never consider the probable interactions (especially for the biologically active chemicals like pharmaceuticals and pesticides) among the components. Apart from this, the CA and IA models need preliminary experimentation, which limits the application of these modeling to some extent. To get rid of these above-mentioned issues, scientists and regulatory agencies are working together; however, a universally acceptable toxicity protocol for mixture toxicity study is still rare. Therefore, an efficient method for the identification and estimation of interactive effects (synergism and/or antagonism) along with the additive effect of mixture components is necessary for better risk assessment of chemicals in a real-world scenario.

The concept of similarity stands upon a relative analysis of different points with respect to others. In the context of ecotoxicology and drug design, it is necessary to have an objective measure of similarity to compare various chemicals concerning their toxicity or activity. Machine learning algorithms are now gaining sufficient attention among scientific communities for predictive modeling of toxicity and biological activity based on chemical similarity. The hybrid modeling approaches can be very useful in this regard due to their ability to combine the detailed process understanding with machine learning; thus they can provide an opportunity to integrate prior process knowledge with various measurement data for efficient modeling of toxicity of hazardous mixtures. Quantitative structure-activity relationships (QSARs), read-across predictions, q-RASAR, and ML algorithms are the most widely used similarity-based computational approaches which can efficiently predict the toxicity of individual chemicals by relative similarity analysis with respect to the chemical structures and properties; hence the use of these *in silico* methods is widely accepted and encouraged by the concerning authorities like the Organization of Economic Co-operation and Development (OECD), European Chemical Agency (ECHA), United States Environmental Protection Agency (US-EPA), and regulations like Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) etc.. Therefore, these methods can be a good alternative for mixture toxicity assessment, which can give predictions based on various structural features of the mixture components. Various research groups have applied QSAR for the mixture toxicity assessment in the recent past and successfully proposed numerous mathematical models for different environmental endpoints. QSAR can also help to identify the interactions among structural features which may be essential for the toxicological effect (biological properties in general). It should be noted here that the QSAR modeling of mixtures is not as simple as the individual toxicants; thus, various additional steps should be followed for successful modeling. Among the key steps, mixture descriptor calculation, and validation of the developed model are the most important. There are very few approaches in the literature, where the cross-validation strategies have been properly used for the mixture toxicity predictions, and hence the chance of overestimation of quality metrics is persisting. Therefore, to study the mixture toxicity of chemical mixtures by in silico models rigorously, more attempts should be made by following the appropriate procedures and without compromising the guidelines of QSAR model development for single chemicals given by OECD.

The approach of the dissertation was to work in the domain of mixture toxicity prediction and risk assessment and add some value in it. Therefore, the objectives of the dissertation were (i) collection of experimental toxicity data of various chemical mixtures against different environmental endpoints, (ii) the development of QSAR models for predicting toxicity of different classes of chemical mixtures (pharmaceuticals, industrial chemicals, plant protection products, etc.), (iii) rigorous validation of the developed models in accordance with the OECD guidelines, (iv) identification of various structural features, responsible for the toxicity of chemical mixtures, (iv) make use of different other machine learning-based new approach methodologies such as read-across and q-RASAR for enhancing the performance of predictive model further.

The title of the performed in silico studies in this thesis work is mentioned below:

Study 1: Prediction of aquatic toxicity of chemical mixtures by the QSAR approach using 2D structural descriptors

Study 2: Application of cross-validation strategies to avoid overestimation of performance of 2D-QSAR models for the prediction of aquatic toxicity of chemical mixtures

Study 3: Chemical similarity and machine learning-based approaches for the prediction of aquatic toxicity of binary and multicomponent pharmaceutical and pesticide mixtures against *Aliivibrio fischeri*

Study 4: Machine learning-based q-RASAR modeling to predict acute contact toxicity of binary organic pesticide mixtures in honey bees

Study 5: "Data fusion" quantitative read-across structure-activity-activity relationships (q-RASAARs) for the prediction of toxicities of binary and ternary mixtures towards three bacterial species

Study 6: Predictive binary mixture toxicity modeling of fluoroquinolone (FQs) and the projection of toxicity of hypothetical binary FQ mixtures: A machine learning approach

The entire work has been presented under the following chapters in this dissertation:

Chapter 1: Introduction

Chapter 2: Present Work

Chapter 3: Materials and Methods

Chapter 4: Result and Discussion

Chapter 5: Conclusion

Reference

Appendix: Reprints

The overall background knowledge of the mixture toxicity assessment, the computational approaches, and the previous works done in this area of research have been discussed under the heading "***Introduction***". Additionally, information on a brief history of QSAR, the basic principles of in-silico approaches that have been employed, and the validation protocol of in-silico models have also been provided. Chapter "***Present Work***" briefly discusses the performed research works along with the collected datasets. In the "***Materials and Methods***" section, the methodology of the entire work done has been thoroughly discussed. Schematic representations of all studies and the important numerical formula used for calculations have been provided in this section to avoid ambiguities. The results of all studies performed have been discussed in the "***Results and Discussions***" in detail. An overall "***Conclusion***" has been provided in this dissertation followed by the "***References***". Most of the studies from this dissertation work have been published in ***reputed international journals*** and also presented at different national and international conferences. Those published works are thus also included in the ultimate section of this dissertation under section "***Reprints***". However, the research work performed and presented in the limited scope of this dissertation constitutes only a small portion in the broad spectram of work in this area of research. Various interesting aspects of research that have arisen in this dissertation, can be further investigated more appropriately in the future study.

In any scientific exploration, success becomes embedded in the explanation of results obtained and the conclusion drawn from it. The results explain the hidden or unexplored scientific facts and the conclusion briefly discusses the new/unexplored information. Access to this new knowledge enables us to gain a deeper understanding of a specific domain of research. On top

of that, results and descriptions are always helpful in clarifying concepts that have already been established and also help to pave the future avenues of research. In the present thesis, the development of in silico studies has been discussed giving special emphasis on the mixtures' toxicity prediction against several ecologically relevant endpoints. Toxicity data of various classes of chemical mixtures i.e., organic chemicals, pharmaceuticals, plant protection products, and different industrial chemicals have been employed. The structural and toxicity data have been curated and then used for the modeling. QSAR, q-RASAR, and ML models have been developed for better and eco-friendly environmental risk assessment of chemical mixtures. Read-across prediction methods have also been employed for the same purpose. The OECD principles of QSAR model development have been followed during the study to avoid any ambiguity in the results obtained. The important structural features (specific for each dataset) have been identified that are responsible for mixture toxicity. All the models have been strictly validated using conventional internal and external validations. Additionally, mixtures-out, compounds-out, and everything-out cross-validations have been employed to avoid overestimation of validation metrics (associated with mixture toxicity modeling). To check the reliability of external validation, the domains of applicability of the developed models have been carried out which also provides the knowledge about the chemical space of respective models. An effort of data gap filling has been made in this thesis where the toxicity of different hypothetical and/or real untested mixtures have been predicted against corresponding ecological species.