

**QUANTITATIVE STRUCTURE – PROPERTY RELATIONSHIP
(QSPR) MODELING OF SELECTED PROPERTY AND
TOXICITY ENDPOINTS OF NANOMATERIALS**

*Thesis submitted in partial fulfillment for the requirements of the Degree of
MASTER OF PHARMACY
Faculty of Engineering and Technology*

Thesis submitted by

JOYITA ROY

B. PHARM

Registration No: **140842 of 2017-2018**

Examination Roll No: **M4PHA19019**

Under the Guidance of

DR. KUNAL ROY

Professor

Drug Theoretics & Cheminformatics Laboratory
Division of Medicinal and Pharmaceutical Chemistry

Department of Pharmaceutical Technology

Jadavpur University

Kolkata – 700 032

India

2019

DECLARATION OF ORIGINALITY AND COMPLIANCE OF ACADEMIC ETHICS

I hereby declare that this thesis contains literature survey and original research as part of my work on “Quantitative Structure – Property Relationship (QSPR) modeling of selected property and toxicity endpoints of nanomaterials”.

All information in this document have been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

NAME: **JOYITA ROY**

EXAMINATION ROLL NUMBER: **M4PHA19019**

REGISTRATION NUMBER: **140842 of 2018-2019**

THESIS TITLE: **“Quantitative Structure–Property Relationship (QSPR) modeling of selected property and toxicity endpoints of nanomaterials”**

SIGNATURE WITH DATE:

CERTIFICATE

**Department of Pharmaceutical Technology
Jadavpur University
Kolkata - 700 032**

This is to certify that **Ms. JOYITA ROY**, B. Pharm. (MAKAUT), has carried out the research work on the subject entitled **“Quantitative Structure–Property Relationship (QSPR) modeling of selected property and toxicity endpoints of nanomaterials”** under my supervision in Drug Theoretics & Cheminformatics Laboratory in the Department of Pharmaceutical Technology of this university. She has incorporated her findings into this thesis of the same title, being submitted by her, in partial fulfillment of the requirements for the degree of Master of Pharmacy of Jadavpur University. She has carried out this research work independently and with proper care and attention to my entire satisfaction.

Dr. Kunal Roy

Professor,
Drug Theoretics and Cheminformatics
Laboratory,
Department of Pharmaceutical Technology,
Jadavpur University,
Kolkata-700 032

(Prof. Pulok Kumar Mukherjee)
Head, Dept. of Pharmaceutical Technology,
Jadavpur University, Kolkata

(Prof. Chiranjib Bhattacharjee)
Dean, Faculty of Engineering and Technology
Jadavpur University, Kolkata

Acknowledgements

*I deem it a pleasure and privilege to work under the guidance of **Dr. Kunal Roy**, Professor, Drug Theoretics & Cheminformatics Laboratory, Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology, Jadavpur University, Kolkata-32. I express my deep gratitude and regards to my revered mentor for suggesting the subject of this thesis and rendering me his thoughtful suggestions and rational approaches to this thesis work. I am greatly indebted to Dr. Kunal Roy for his valuable guidance throughout the work that enabled me to complete the work. With a deep sense of thankfulness and sincerity, I acknowledge the continuous encouragement, perpetual assistance and co-operation from my seniors Probir Kr. Ojha, Priyanka De, Khan Kabiruddin, Vinay Kumar, Gopal krishan and Mohsin Khan Pathan. Their constant support and helpful suggestions have tended me have tended me to accomplish this work in time. I would like to express my special thanks to my friend Kazi Amirul Hossain and my juniors Sapna Pandey and Arnab Seth who all have extended their helping hands and friendly cooperation all through my work.*

I am indeed glad to convey cordial thanks to all my friends specially Kuntal Saha Chowdhury, Sulekha Ghosh, Reshma Prasad and Ajeya Samanta. I am thankful to the authority of Jadavpur University and Head of the Department Prof. Dr. Pulok Kumar Mukherjee for providing all the facilities to carry out this work.

A word of thanks to all those people associated with this work directly or indirectly whose names I have been unable to mention here. Finally, I would like to thank my parents Mr. Hiralal Roy and Mrs. Debika Roy for all the love and inspirations without which my dissertation work would remain incomplete.

Joyita Roy

Examination Roll No: M4PHA19019

Department of Pharmaceutical Technology,

Jadavpur University,

Kolkata-700032

Preface

This dissertation is presented for the partial fulfillment of the degree of Master of Pharmacy in Pharmaceutical Technology. The work presented in this dissertation is spread over a span of two years. The present work has been investigated through the development of predictive *in silico* models on toxicity and property on nanomaterials. Now-a-days, different *in silico* techniques are being routinely applied for predicting toxicity and property of different nanomaterials as a rational alternative technique to animal testing. The use of statistical models to predict biological and physicochemical properties started with linear regression models developed by Hansch in the 1960s. Since then the appearance of computer-aided drug design studies, the term “Quantitative Structure-Activity Relationships (QSARs)” has become one of the most popular terms in medicinal, environmental and synthetic chemistry. Development of predictive quantitative structure-property relationship (QSPR) models allows assessment of toxicological hazard and property of nanomaterials from the chemical information derived using descriptors. Here, we have especially considered some property and toxicity endpoints of the nanomaterials.

Nanotechnology holds incredible promise benefits in a wide range of industrial applications, cosmetics, food safety, environmental science to material science, information and communication technology, transportation, and many others. The rapid increase in engineered nanomaterials in recent years is because of its unique Magnetic, electronic and optical properties. The potential benefits to society include stronger, lighter, more durable materials, remote sensing and tracking devices related to food quality and spoilage, improved systems to control, prevent, and remediate pollution problems or cost effective development and use of renewable energy sources. Along with the advantages of nanotechnology there is some problem or risk related to it. According to recent studies, nanomaterials may endanger human health through the potential induction of cytogenetic, genotoxic, mutagenic and ecotoxic effects. It is obvious that animal testing cannot alone suffice the need for the entire assessment of those nanomaterials. So, alternative method should come in. Hence, assessment of properties of the existing as well as new nanomaterial is necessary.

The QSPR analysis has got its important application in the field of drug discovery and virtual screening for designing of more potent nanoparticle with specific activity and reduced toxic adverse effects. Different regulatory agencies support the use of QSPRs method. The concept of correlation of chemical structure and the property (including activity, toxicity, and physicochemical property) gave rise to the development of QSPR analysis for addressing a wide number of endpoints. Such studies complement the 3Rs (replacement, refinement and reduction of animals in research) minimizing animal testing. The present dissertation attempts to explore the QSPR approach by using only 2D descriptors and periodic table descriptors towards various endpoints of nanomaterials. It has been found that 2D descriptors have been very interestingly successful in modeling property of chemical compounds towards CNTs (carbon nanotubes). Application two-dimensional representation of the molecules in the model requires relatively

lesser computational time than the other ones. Similarly, periodic table descriptors are simple and uses less computational experiences, they could be taken easily from the periodic tables without vigorous calculation. The QSPR models reported in this dissertation have been judged in the light of the five point seminal guidelines proposed by the Organization for Economic Co-operation and Development (OECD) involving data uniformity, methodological explicitness, and domain of applicability, statistical reliability, and mechanistic interpretability. The models developed have showed acceptable statistical significance. We have developed predictive QSPR models on different endpoint towards nanoparticles. The models developed were validated rigorously based on internal and external validation strategies. The following analyses have been performed in this dissertation.

Study 1: Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)

Study 2: Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors

The accomplished work has been presented in this dissertation under the following sections:

Chapter 1: Introduction

Chapter 2: Present Work

Chapter 3: Materials and Methods

Chapter 4: Results and Discussion

Chapter 5: Conclusion

Chapter 6: References

Appendix: Reprints

Abbreviations

Abbreviations	Full forms	Abbreviations	Full forms
AD	Applicability domain	NPs	Nanoparticles
ANN	Artificial neural network	OCs	organic chemicals
CNTs	Carbon nanotubes	OECD	Organisation for Economic Co-operation
<i>df</i>	Degrees of freedom	PCA	Principal Component Analysis
DNA	Deoxyribonucleic Acid	PCR	Principal Component Regression
E_c	values of conduction band energy	PLS	Partial Least Squares
EC ₅₀	50% effective concentration	pm	picometer
ED	Euclidean distance	POPs	Persistent organic pollutants
E_{ca}	electron affinity	PRESS	Predicted residual sum of squares
Eq	Electrochemical Equivalent	QAAR	quantitative activity–activity relationship
ETA	Extended Topochemical Atom	QSAR	Quantitative structure-activity relationship
E_v	valence band energy	QSPR	Quantitative structure-property relationship
H ₂ O ₂	hydrogen peroxide	QSTR	Quantitative structure-toxicity relationship
ICP	intelligent consensus predictor	QTTR	quantitative toxicity–toxicity relationship
kNN	k-nearest neighbors	R_c	covalent radius
LC ₅₀	50% lethal concentraion	REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
LCA	Life cycle assessment	rmsep	Root mean square error in prediction
LDA	Linear discriminant analysis	ROC	Receiver operating characteristics
LFER	Linear free energy relationships	ROS	reactive oxygen species
LOO	Leave-one-out	RTO	regression through origin
LR	Linear Regression	SAR	Structure Activity Relationship

LV	Latent variable	SD	Standard deviation
MAE	Mean absolute error	SDEP	Standard deviation of error of prediction
MLR	Multiple Linear Regression	SEE	Standard error of estimate
mtDNA	mitochondrial DNA	SOM	Kohonen's Self-Organizing Map
MW	Molecular weight	TAU	Topochemically arrived unique
MWNTs	multi walled nanotubes	T_c	Thermal conductivity
NADPH	Nicotinamide adenine dinucleotide phosphate	VOCs	volatile organic compounds
nano-QSTR	Nano-Quantitative Structural Toxicity Relationship	NPs	Nanoparticles

CONTENTS

Chapter	Topic	Page no.
	Acknowledgement	i
	Preface	ii-iii
	Abbreviations	iv-v
1	Introduction	1-26
1.1	What are QSPR /QSAR Modeling?	2-4
1.1.1	The objectives of QSAR	4
1.1.2	Importance of QSAR/QSPR	5
1.1.3	Descriptors	5-9
1.1.4	Classification of QSAR methodologies	9-10
1.1.5	Development of quantitative models over the past century	10-14
1.1.6	The methodology of QSAR	14-18
1.1.7	Application of QSAR/QSPR	18-19
1.2	Nanomaterials	19-20
1.2.1	Types of nanomaterials	20-22
1.2.2	Applications	22
1.2.3	Health and safety	23
1.3	Role of predictive QSPR models on nanomaterial/nanoparticles	24-26
2	Present work	27-31
2.1	Study 1: Dataset 1	28-30
2.2	Study 2: Dataset 2	30-31
3	Materials and method	32-81
3.1	Details of datasets consisting chemical structures along with their activity or toxicity data	32
3.1.1	Dataset I (study 1)	32-43
3.1.2	Dataset II (study 2)	43-45
3.2	General description of methods implemented for developing QSAR models	46
3.2.1	Descriptor calculation	46-59
3.2.2	Pretreatment of descriptors: thinning of the pool	59
3.2.3	Division of the dataset: selection of the training set and test set	60-61

	3.2.4	Selection of feature	61-62
	3.2.5	Employed statistical modeling techniques	62-64
	3.2.6	Computation of different statistical metrics for assessing model quality	64-75
	3.2.7	Software packages employed in the study	76
3.3		Study wise specific description of methodologies utilized in each study	76
	3.3.1	Study 1: Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)	76-78
	3.3.2	Study 2: Risk assessment of heterogeneous TiO ₂ -based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors	79-81
4		Results and discussions	82-114
	4.1	. Study 1: Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)	82-100
	4.1.1	The descriptors related to hydrogen bonding interactions	91-93
	4.1.2	The descriptors related to hydrophobic interactions	93-96
	4.1.3	The descriptors related to π - π interactions	96-98
	4.1.4	The descriptors related to electrostatic interactions	98-100
	4.2	Study 2: Risk assessment of heterogeneous TiO ₂ -based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors	100-111
	4.2.1	Electrochemical Equivalent (Eq)	103-104
	4.2.2	2nd ionization potential ($2\chi_{pi}$)	105-107
	4.2.3	Covalent radius (Rc)	107-108
	4.2.4	Amount of Ag (Agamt)	108-109
	4.2.5	Thermal conductivity (Tc)	109-111
	4.2.6	PLS model	111-114
5		Conclusion	115-119
	5.1	Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)	115-118
	5.2	Risk assessment of heterogeneous TiO ₂ -based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors	118-119
6		References	120-134
7		Appendix - Reprints	

CHAPTER - 1

Introduction

1. INTRODUCTION

One of the greatest scientific and technical achievements at the end of the 20th century is the creation of nanomaterials and nanotechnology. As per many experts, scientific and technological advance in the present century is determined by the achievements associated with nanomaterials, nanotechnology, and their application in various fields of natural science and technology and in other fields of human activity. A diverse array of consumer goods are produced at the nanoscale, which is usually defined by at least one diameter of the considered species to be of 100 nm or less. Nanomaterials due to their unique magnetic, electrical, optical, thermal, and chemical properties are applied in huge spectrum and their applications are increasing day by day [1], this “nano enthusiasm” in present century however, should go hand in hand with care and precaution. Some recent publications reported evident toxicity of selected nanoparticles and highlighted their potential risk connected with the development of nano engineering. So, a thorough understanding and knowledge of the relationship between the physicochemical properties and the behavior of nanomaterials in biological systems is mandatory for designing safe and efficacious products. Since, experimental evaluation of the safety of nanomaterials is expensive and time-consuming, computational methods have been found to be efficient alternatives for predicting the potential toxicity and environmental impact of new nanomaterials before mass production.

Quantitative structure-activity relationship (QSAR) methods provide one option for establishing such relationships. A QSAR is a statistical model that relates a set of structural or property descriptors of a chemical compound to its biological activity. The QSAR methodology is well known and extensively applied in the areas of drug discovery and chemical toxicity modeling for guiding the experimental design of chemical compounds, and its growing importance for providing key information is reflected in a number of regulatory frameworks (e.g., Registration Evaluation Authorization and Restriction of Chemical Substances (REACH)) where QSAR approaches are considered acceptable methods under certain conditions for filling in knowledge gaps for untested chemicals.

1.1. What are QSAR /QSPR Modeling?

QSAR modeling on a set of structurally related chemicals refers to the development of a mathematical correlation between a chemical response and quantitative chemical attributes defining the features of the analyzed molecules. Such study attempts to establish a mathematical formalism between the behavior of a chemical, i.e., chemical response and a set of quantitative chemical attributes which may be extracted from the chemical structures using suitable experimental or theoretical means. The naming of the study depends on the end points (i.e nature of response) which can be classified into three class namely quantitative structure-property/activity/toxicity/ relationship (QSPR/QSAR/QSTR) studies taking into consideration the biological activity and toxicological data respectively [2]. On the other hand, QSPR, i.e., quantitative structure–property relationship modeling, can be employed to designate all such related techniques as any type of biological and toxicological as well as physicochemical behavior may be considered as the ‘property’ of a given chemical. Therefore, a basic mathematical formalism can be develop and represented as follows in **Eq.1.1**.

$$\text{Biological activity} = f(\text{Chemical attributes}) = f(\text{Structure, Property}) \quad (1.1)$$

The phrase “*Chemical attributes*” denotes the features that define the behavior manifestation or in other words they are the fundamental knowledge of the chemicals which control the behavioral response under investigation. The behavioral manifestation can be well explained by its physiological properties which represent the intrinsic molecular nature. The QSPR [3] study deals with the molecular features governing their physicochemical properties. The descriptors measure properties of the molecules which include their hydrophobic, steric and electronic features in addition to the various structural patterns while the QSTR [4] technique determines the structural attributes of the molecules responsible for their toxicity profile. The chemical attributes usually describe the informations obtained directly from the structure and their physiological information is obtained from experimental techniques leading to the respective expression as represented in **Eq. 1.2**.

$$\text{Response} = f(\text{Chemical attributes}) = f(\text{Structure, physiological Property}) \quad (1.2)$$

Considering the employment of a series of chemical information in presence and absence of physicochemical features, the QSAR equation for a specific response can be mathematically stated as follows in **Eq.1.3**.

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n \quad (1.3)$$

Since we are talking in terms of a mathematical correlation, such equations are better explained in terms of variables. Here, 'y' represents respond being modeled i.e. activity, toxicity, property while $X_1, X_2, X_3 \dots X_n$ represent the independent variables denoting the physiological properties in terms of numerical quantities and $a_0, a_1, a_2, a_3 \dots a_n$ stand for the contribution of individual descriptors with a_0 as constant term. The physiological property not only can be used as response or dependent variable giving a QSPR (Quantitative structure-property relationship) but also as independent variable or predicted variables. Such studies are termed as quantitative activity-activity relationship (QAAR) or quantitative toxicity-toxicity relationship (QTTR) or quantitative property-property relationship (QPPR) modeling, as appropriate. The QSAR analysis is principally aimed at quantification of chemical information followed by developing a suitable interpretative relationship addressing a given response. Thus, mathematics here serves as a tool for deriving a suitable relationship which is then exploited as per the requirement of the designer [5]. A QSAR study encompasses avenue of chemistry and physics accounting for intrinsic molecular nature, mathematics and statistics for modeling and calculation, and biology to encompass the involved biochemical interaction. The QSAR study can be visualized to comprise three simple steps, namely (a) data preparation, (b) data processing, and (c) data interpretation for a set of chemicals. The quantitative data are obtained from two major components, namely the response or endpoint to be addressed and the predictor or independent variables (i.e., X variables) defining the chemical attributes. The first step, i.e., the preparation of data involves arrangement and conversion of the data in a suitable form. The response data for various biological and toxicological endpoints are usually obtained in two forms, namely 'dose-fixed response' pattern where the dose or concentration of a chemical required to produce a desired fixed response is measured and 'response-fixed dose' pattern in which the response elicited by a chemical at a fixed dose (concentration) is opted for. The quantitative data for the predictor variables are obtained from experimental observations usually comprising different physicochemical measures as well as theoretical calculations. Since response values for these analyses being obtained from multiple assays at different dose or concentration levels of chemicals, these (i.e., doses required to elicit a fixed response) are preferably used as the independent variable (Y) in QSAR studies. Hence, a model can be developed from the

information of varying concentrations of chemicals required to exhibit a fixed biological (or toxicological) response. The quantitative data for the predictor variables are obtained from experimental observations that comprise of different physicochemical measures as well as theoretical calculations. Finally, a data matrix is prepared in which rows present different chemicals in the data set while the response variable and several independent predictor variables are presented in columns. QSAR studies employ computation of several statistical measures and metrics to characterize the quality, stability, and validation of the models. The final operation, i.e., the interpretation of the developed model, is very crucial and it requires a thorough knowledge on the biochemical aspects of the molecules toward the response being modeled. It might be noted that QSAR modeling eventually attempts to establish a chemical basis for specific phenomena such as activity, property, or toxicity by the development of a suitable correlation equation or model.

1.1.1. Objectives of QSAR/QSPR

The principle objectives of QSAR analysis are:

1. Prediction of new analogs of compounds with better property
2. Better understanding and exploration of the modes of actions
3. Optimization of the lead compound with decreased toxicity [6].
4. Reduction of the cost, time and manpower requirement by developing of more effective compounds using a scientifically less exhaustive approach.

To achieve the aforementioned objectives, it is necessary to have a detailed knowledge on the following aspects:

- (i) Detailed knowledge of the mode of action of the molecules.
- (ii) Various factors controlling the experimental condition of the molecules.
- (iii) A thorough examination of molecular structures and their properties. Quantitative structure-activity relationship is an interdisciplinary study of chemistry, biology, and statistics. By the prediction of the essential structural requirements needed for obtaining a molecule with optimized activity/toxicity/property, QSAR analysis provides a good platform for the

synthesis of relatively lesser number of chemicals with improved activity toxicity/property of interest.

1.1.2. Importance of QSAR/QSPR

Although the development of predictive QSAR/QSPR/QSTR models appears to be a simple task, it has got enormous applications in serving the need of scientific fraternity. Sometimes different chemical compounds elicit different response profile but it is rather astonishing that even the same chemical can elicit different biological actions. Hence, it become crucial for determining the chemical features responsible for the behavioral changes. QSAR techniques provide several advantages in terms of model predictivity and utilization of limited experimental resources, employing less computational time. QSAR can be applied in costly areas like drug discovery and development where valuable rational designing would be provided with minimal cost involvement. It may be noted that QSAR helps in achieving efficient, effective, safe, and environmentally benign chemicals and processes thereof and thereby facilitates a ‘sustainable chemical’ process [5].

1.1.3. Descriptors

Molecular descriptors are terms that characterize specific information about a studied molecule. They are the “numerical values associated with the chemical constitution for correlation of chemical structure with various physical properties, chemical reactivity, or biological activity” [7-8]. In other words, the modeled response (activity/property/toxicity) is represented as a function of quantitative values of structural features or properties that are termed as descriptors for a QSAR model as shown in **Eq. 1.4**. Chemo informatics methods depend on the generation of chemical reference spaces into which new chemical entities are predictable by the developed QSAR model. The definition of chemical spaces significantly depends on the use of computational descriptors of studied molecular structure, physical or chemical properties, or specific features.

$$\text{Response (activity/property/toxicity)} = f(\text{information in the form of chemical structure or property}) = f(\text{descriptors}) \quad (1.4)$$

The type of descriptors used and the extent to which they can encode the structural features of the molecules that are correlated to the response are critical determinants of the quality of any QSAR model. The descriptors may be physicochemical (hydrophobic, steric, or electronic), structural (based on frequency of occurrence of a substructure), topological, electronic (based on molecular orbital calculations), geometric (based on a molecular surface area calculation), or simple indicator parameters (dummy variables).

It is interesting to point out that the efficacy of a descriptor can rely heavily on the problem being considered. More precisely, certain end points may need to take into account exact molecular features. The best possible features that make a descriptor ideal for the construction of a QSAR model are summarized here:

1. A descriptor must be correlated with the structural features for a specific end point and show negligible correlation with other descriptors.
2. A descriptor should be applicable to a broad class of compounds.
3. A descriptor that can be calculated rapidly and does not depend on experimental properties can be considered more suitable than one that is computationally exhaustive and relies heavily on experimental results.
4. A descriptor should generate dissimilar values for structurally different molecules, even if the structural differences are small. This means that the descriptor should show minimal degeneracy. In addition to degeneracy, a descriptor should be continuous. It signifies that small structural changes should lead to small changes in the value of the descriptor.
5. It is always important that the descriptor has some form of physical interpretability to encode the query features of the studied molecules.
6. Another significant aspect is the ability to map descriptor values back to the structure for visualization purposes [9]. These visualizations are sensible only when descriptor values can be associated to structural features.

1.1.3.1. Types of Descriptors

Descriptors can be of different types depending on the method of their computation or determination: physicochemical (hydrophobic, steric, or electronic), structural (frequency of

occurrence of a substructure), topological, electronic (molecular orbital calculations), geometric (molecular surface area calculation), or simple indicator parameters (dummy variables).

In a broader perspective, descriptors (specifically, physicochemical descriptors) can be classified into two major groups: (1) substituent constants and (2) whole molecular descriptors [10-11].

1) **Substituent constants:** physicochemical descriptors which are designed on the basis of factors, which govern the physicochemical properties of chemical entities.

2) **Whole molecular descriptors:** expansion of substituent constants approach.

Descriptors can be also classified based on the dimension as

Sl. No.	Dimension of descriptors	Parameters
1	0D-descriptors	Constitutional indices, molecular property, atom and bond count
2	1D-descriptors	Fragment counts, fingerprints
3	2D-descriptors	Topological parameters, structural parameters, physicochemical parameters including thermodynamic descriptors
4	3D-descriptors	Electronic parameters, spatial parameters, molecular shape analysis parameters, molecular field analysis parameters and receptor surface analysis parameters

1.1.3.1.1. 2D-Descriptor

I) Physiological descriptors

These descriptors are derived from some physiological experimentation findings and have a connection with physiological properties. Due to change of the physiological property the adsorption, distribution and excretion will also change. The important physicochemical factors affecting bioactivity of drugs and chemical include hydrophobicity, electronic, and steric character of the whole molecules and also the substituent present in the molecules [12].

II) Indicator variables

They are used in QSAR due to their simplicity, they are substructure based descriptors. They differ to each other by substructure existing in one set but not the other can be studied entirely by using only indicator variables. This could be used when the two sets are identical to each other in all respect (limitation) expect the substructure being coded with the indicator variables.

III) Topological descriptors

Topological descriptors are calculated based on the graphical representation of molecules and so they neither require estimation of any physicochemical properties nor need the rigorous calculations that are involved in the derivation of the quantum chemical descriptors. The structural representation depends on the 2D- graphical topology which indicates the position of the atoms and the connected bonds. It is formed on the graph theory where the molecule structure vertices represents the atoms and edges as the covalent bonds [13].

1.1.3.1.2. 3D-Descriptors

I) Electronic Parameters

Electronic descriptors are defined in terms of atomic charges and are used to describe electronic aspects both of the whole molecule and of particular regions, such as atoms, bonds, and molecular fragments. Electrical charges in the molecule are the driving force of electrostatic interactions, and it is well known that local electron densities or charges play a fundamental role in many chemical reactions and physicochemical properties [10].

II) Spatial Parameters

Spatial parameters comprise a series of descriptors calculated based on the spatial arrangement of the molecules and the surface occupied by them [10].

1.1.4. Classification of QSAR methodologies

1.1.4.1. Based on dimensionality

Dimension	Method	Reference
0D-QSAR	Models are based on descriptors involving molecular formula like molecular weight etc	[14]
1D-QSAR	Models are based on the simplex representation of molecular structure (SiRMS) approach.	[15]
2D-QSAR	Activity is correlated with physicochemical and structural patterns (connectivity, topology etc.) of the molecules without consideration of an explicit 3D representation of these properties.	[10]
3D-QSAR	Activity is correlated with three-dimensional structure of the ligands.	[16]
4D-QSAR	Ligands are represented as an ensemble of configurations	[17]
5D-QSAR	As 4D-QSAR + explicit representation of different induced-fit models	[18]
6D-QSAR	As 5D-QSAR + simultaneous consideration of different solvation models	[19]

1.1.4.2. Based on the type of chemometric methods used

Sometimes QSAR methods are also classified into following two categories, such as

- Linear methods Linear regression (LR), multiple linear regression (MLR), partial least squares (PLS), and principal component analysis/regression (PCA/ PCR) and
- Nonlinear methods Artificial neural networks (ANN), k-nearest neighbors (kNN), and Bayesian neural nets.

1.1.5. Development of quantitative models over the past century

QSAR methods originated way back in the nineteenth century. The order of evolution of different methodologies focusing more precisely on the essential molecular structural attributes over the decades.

1.1.5.1 De Novo Models

De novo QSAR models are the mathematical models which do not require computation of any descriptors encoding chemical information on molecular structure. Indicator parameters (having a binary value 0 or 1) representing presence or absence of a group at a particular position are used for development of the models.

i) Hansch's Method

In 1962 Hansch et.al correlated the plant growth regulatory activity of phenoxyacetic acids to Hammett constants and partition coefficients [20]. Two years later they showed that the biological activity could be correlated linearly with free-energy related terms [21]. This model was known as Linear Free Energy Relationship (LFER) and later changed to extra thermodynamic approach and expressed by (Eq.1. 5).

$$\log 1/C = a\pi + b\sigma + cEs \dots \dots + constant \quad (1. 5)$$

Where, C = molar concentration of compound to produce defined biological response.

π = the hydrophobic contribution of the substituent and represented by $\log P_X/P_H$.

σ = Hammett electronic descriptor of the substituent [22].

E_S = Taft's steric parameter [23].

a, b, c = appropriate constants.

Later Hansch and co-workers realized that the activity of the hydrophobic drugs decreased after reaching optimum concentration. This was due to entrapment of drugs in the lipid phase of the transport process [21]. Due this reason they introduced a new model (parabolic model) for this type of molecules (**Eq.1. 6**).

$$\log \frac{1}{c} = -k(\log P)^2 + K' \log P + K'' \quad (1. 6)$$

In the above equation the meaning of C is same as previous equation and k, k', k'' are constants which are obtained via the method of least squares.

ii) Free-Wilson model

The Free-Wilson approach is truly a structure activity-based methodology because it includes the contributions made by various structural fragments to the overall biological activity [24-25]. It is represented by the (**Eq.1. 7**).

$$BA_i = \sum a_j X_{ij} + \mu \quad (1. 7)$$

Here, BA = biological activity,

X_j = j th substituent, which carries a value 1 if present, 0 if absent,

a_j = contribution of the j th substituent to biological,

μ = overall average biological activity.

This de novo approach assumes that effects of substituents are additive and constant [26]. This approach does not need of physicochemical constant. However, there are certain limitations. Large number of variables is required to describe a smaller number of compounds together with a large number of molecules with varying substituents. Besides, these intra-molecular interactions are not handled well. The constant term (μ) is an overall average of the biological activity of all the compounds used to develop the model.

iii) Fujita-Ban Model

Fujita and Ban [27] modified the approaches of Free-Wilson model. In this model the biological activity is expressed in logarithmic scale. It is a free energy-related term and additive in nature, represented by **Eq.1.8**.

$$\log\left(\frac{A}{A_0}\right) = \sum G_i X_i \quad (1.8)$$

'A' and 'A₀' are the magnitude of the activity of substituted and unsubstituted compounds respectively. G_i is the log activity contribution or the log activity enhancement factor of the ith substituent relative to that of H and X_i is the parameter which takes a value 1 or 0 according to the presence or absence of the ith substituent.

Advantages of this model are:

- i. The structural matrix does not need to be transformed,
- ii. No restriction equation is necessary,
- iii. The group contribution at each position is based on the parent compound (i.e., H),
- iv. The constant term (m) is calculated by the least square method and is the theoretically predicted value for the unsubstituted compound.

The addition or omission of a compound does not affect markedly the value of group contributions.

1.1.5.2. Property-based QSAR

i) LFER Approach of Hansch

Physiological properties as descriptors have been widely used in the QSAR/QSPR modeling studies. Hydrophobic, electronic, and steric are the three main physiological properties used for modeling (either whole molecule or substituents). It was originally promoted through linear free energy-related (LFER) approach. The LFER approach of Hansch using physicochemical descriptors and substituent constants has its origin in the work of Hammett

[28] in physical organic chemistry. Hammett defined an electronic substituent constant σ for the hydrolysis rates of benzoic acid derivatives in the following expression (**Eq.1.9**).

$$\log\left(\frac{K_x}{K_H}\right) = \rho\sigma \quad (1.9)$$

Here, K_x and K_H stands for the equilibrium constants for the reaction of the substituted and unsubstituted benzoic acid, ρ is a constant dependent on type and conditions of the reaction as well as the nature of compounds, σ is an electronic substituent constant depending on its nature and position of the substituent. The equation can also be written as **Eq.1.10**.

$$\log K_x = \rho\sigma + \log K_H \quad (1.10)$$

Note that Hammett σ is applicable for Meta- and para-aromatic substituents. In analogy to the Hammett σ equation, Hansch and Fujita [29] introduced another substituent constant π in the following manner (**Eq.1.11**)

$$\pi_X = \log\left(\frac{P_X}{P_H}\right) \quad (1.11)$$

In the above equation, π_X is the hydrophobic substituent constant of substituent X, while P_X and P_H are (n-octanol–water) partition coefficients of substituted and unsubstituted compounds.

Hansch observed a parabolic dependence of the biological activity on the hydrophobicity or hydrophobicity constant (**Eq.1.12** and **Eq.1.13**).

$$\log \frac{1}{C} = a\pi - b\pi^2 + C \quad (1.12)$$

$$\log \frac{1}{C} = a\log P - b\log P^2 + C \quad (1.13)$$

Here, it uses the hydrophobicity term $\log P$ for the whole molecules.

On using both electronic and hydrophobic substituent constant terms, a generalized expression of Hansch equation can be shown as follows in **Eq. 1.14**.

$$\log \frac{1}{c} = K_1\pi - K_1\pi^2 + K_1\sigma + K_4 \quad (1.14)$$

Additionally, when E_s (steric) is needed it may be added as following as represented in **1.15**.

$$\log \frac{1}{c} = K_1\pi - K_1\pi^2 + K_1\sigma + K_4E_s + K_5 \quad (1.15)$$

All descriptors appearing in the final model should have statistically significant regression coefficients; otherwise, such terms should be omitted. In selecting the physicochemical parameters to be used in the QSAR models, one should check the possibility of intercorrelation among various pairs of substituent constants.

ii) Mixed approach

The Hansch approach and the Fujita–Ban model can be combined to a mixed approach. If for one definite region of the molecule, a Hansch correlation can be obtained for the substituents, while substituents in another position of the molecule must be treated by Free–Wilson analysis (using indicator variables for the presence or absence of substituents at particular positions), the Fujita Ban model and the Hansch approach (**Eq.1.16**) can be combined to a mixed approach as given in the following expression [30].

$$\log \frac{1}{c} = K_1\pi + K_1\sigma + \Sigma G_i X_i + C \quad (1.16)$$

Here, $K_1\pi + K_1\sigma$ is the Hansch part for the substituents Y_j and $\Sigma G_i X_i$ is the (modified) Free–Wilson part for the substituents X_i while c is the theoretically predicted activity value of the unsubstituted parent compound ($X = Y = H$) or of an arbitrarily chosen reference compound.

1.1.6. The methodology of QSAR

Development of predictive QSAR models comprises several steps namely i) data preparation ii) data analysis iii) data validation and iv) data interpretation where the “data” refers to the response and predictor variables. The steps are briefly discussed below.

i) Data preparation:

- First the biological/toxicological/physiological response is converted to the required unit and maintaining data uniformity.
- Second, drawing of the chemical structures using suitable drawing software like ChemDraw, MarvinSketch, ChemSketch etc. The chemical structures can also be collected from public databases such as PubChem (<https://pubchem.ncbi.nlm.nih.gov/search/index.html>), NIST Chemistry. The configuration should be checked before using the structures.
- Depending on the purpose of modeling they should be perform energy minimization operation and conformational analysis.
- The file containing the structure is subjected to software for descriptor (theoretical predictor variables) calculation. Initially the descriptors are subjected to data pretreatment to remove the constants. Different application can be used for the descriptor calculation i.e different class of descriptors.
- The descriptors containing different variables and single column of response (activity/ property/ toxicity) are compiled in a single worksheet which is the QSAR data matrix. An additional column denoting the name of the chemicals can be added for quick identification of any compound.

ii) The data analysis:

This component consists of feature selection, dataset division and model development.

- Selection of features refers to the identification of the important predictor variables suitable for developing correlation with the response variable. Usually various feature selection tools are coupled with one or more model development techniques under the same interface so that the user can select the best possible predictor variables and develop models using them at the same time. Many applications are capable of generating hundreds or thousands of different molecular descriptors. Some of the feature selection tools employed in chemometric modeling studies includes stepwise variable selection, genetic algorithm, best subset selection, variable subset selection, and factor analysis. Typically, only some of them are significantly correlated with the activity. Furthermore, many of the descriptors are intercorrelated. This has negative effects on

several aspects of QSAR analysis. Some statistical methods require that the number of compounds is significantly greater than the number of descriptors. Using large descriptor sets would require large datasets.

- Selection of the training set compounds is significantly important in QSAR analysis [31-32]. In order to ascertain the performance of a predictive model the whole dataset is divided into a training set and a test set based on chemical similarity. The training set (i.e., the equation), while the test set (not used during model development) is used to judge the external predictivity of the model. However, the most rational approach of selection of training set is based on relevant physicochemical descriptors and chemical similarity principle. Higher number of compounds is allotted in the training set as it is used in the development of the model. The process is based on the assumption that a molecule structurally very similar to the training set molecules will be predicted well because the model has captured features that are common to the training set molecules and is able to find them in the new molecule. Care should be taken during selection of the training and test sets such that the test set compounds lie within the structural domain of the training set compounds. Otherwise complete changes in structural similarity of the test set chemicals from those in the training set will result in a poor prediction by the model developed using the training set. The methods for the selection of training and test set are:

- (i) k-means clustering and Kennard-Stone selection
- (ii) Principal component analysis (PCA)
- (iii) Kohonen's Self-Organizing Map (SOM)
- (iv) D-optimal design
- (v) Sphere exclusion
- (vi) Sorted response

Here, the whole data matrix is first sorted using the response column followed by selection of a predefined fraction of compounds into training/ test set from different zones maintaining a pattern e.g., every fourth compound, etc. In the random division approach, compounds are randomly classified into training and test sets following a user defined fraction. Sometimes

combination of response variable based and predictor variable based approach may also be employed e.g., compounds may be assigned into different structurally similar groups using any of the above mentioned techniques followed by selection of compounds into training/test set using the sorted response formalism separately from each group.

- The model development step dictates that the selected best features are to be combined in a single equation employing an explicit formalism. After the calculation of different features, i.e. descriptors, construction of QSAR model is done by using feature mapping procedure also referred to as the parameter estimation problem. The aim is to construct a pure mathematical relationship between the descriptors and the response under investigation. Multiple linear regression (MLR), partial least squares (PLS) etc. are the algorithm employed for the development of quantitative regression based equations while linear discriminant analysis (LDA) yields classification model. The variable selection tools are accompanied by statistical evaluation of the corresponding model developed from the selected variables as stepwise-MLR, GFA-MLR, G/PLS (genetic PLS), PLS-DA (PLS followed by discriminant analysis) etc. finally based on model validation criteria, the best model is chosen for further analysis [32].

iii) Model validation:

Following the development of predictive models, next essential task becomes the determination of its statistical reliability. Since the objective of QSAR analysis is not only to develop a model but also to apply it for the prediction of response of untested/ new chemicals, it is necessary to ascertain its stability as well as predictivity. Various statistical metrics are computed to judge the model fitness (R^2 , R_a^2 etc.), internal stability (Q^2_{LOO} , $r_m^2(\text{LOO})$) as well as external predictivity (R^2_{pred} , $r_m^2(\text{test})$) and values above the threshold limits identify model acceptability. It may be noted that by “internal stability” we aim to portray the stability of prediction determined using the training set compounds only, i.e., compounds used for developing the model, while external predictivity refers to the judgment on test set prediction. Some additional metrics can also be employed to judge the overall predictivity e.g., $r_m^2(\text{overall})$. For the validation of discriminant model parameters such as sensitivity, specificity, accuracy, precision, F -value, receiver operating characteristic (ROC) analysis etc. can be employed.

iv) Model interpretation:

Once a QSAR model has been developed and has been considered acceptable from the values of the metrics, the final important part remains with the mechanistic interpretability of the modeled features. Establishing a suitable basis between the chemistry of the compounds and biological/ toxicological action or physicochemical property helps in understanding the mechanism of action involved. Accordingly, by combining the experimental results and observation from the model, one can explicitly explain each step of the process of behavioral manifestation of chemicals. Such knowledge is useful in designing and developing potent analogues.

1.1.7. Application of QSAR/QSPR

QSAR presents a suitable option in the rational monitoring of activity/ property/toxicity of chemicals and hence is useful in a wide variety of applications (**Figure 1**) namely biological activity, predictive toxicity and physicochemical property. Since fine-tuning of the behavioral nature of chemicals gives fruitful results for a significantly large class of chemicals such as

- Pharmaceuticals
- Agrochemicals
- Perfumeries
- Analytical reagents
- Solvents
- Surface modifying agents etc

The chemicals modeled using the QSAR method can be overviewed in three major types, namely

1. Chemicals of health benefits (drugs, pharmaceuticals, food ingredients, etc.),
2. Chemicals involved in industrial/laboratory processes (solvents, reagents, etc.), and
3. The chemicals posing hazardous outcome persistent organic pollutants (POPs), toxins, xenobiotics, volatile organic compounds (VOCs).

Besides modeling biological activity and toxicity endpoints, it may also involve in modeling of ADME which involve in pharmacokinetics profile of drug candidates prior to its synthesis and hence enhancing the efficacy of the designed drug in biological system.

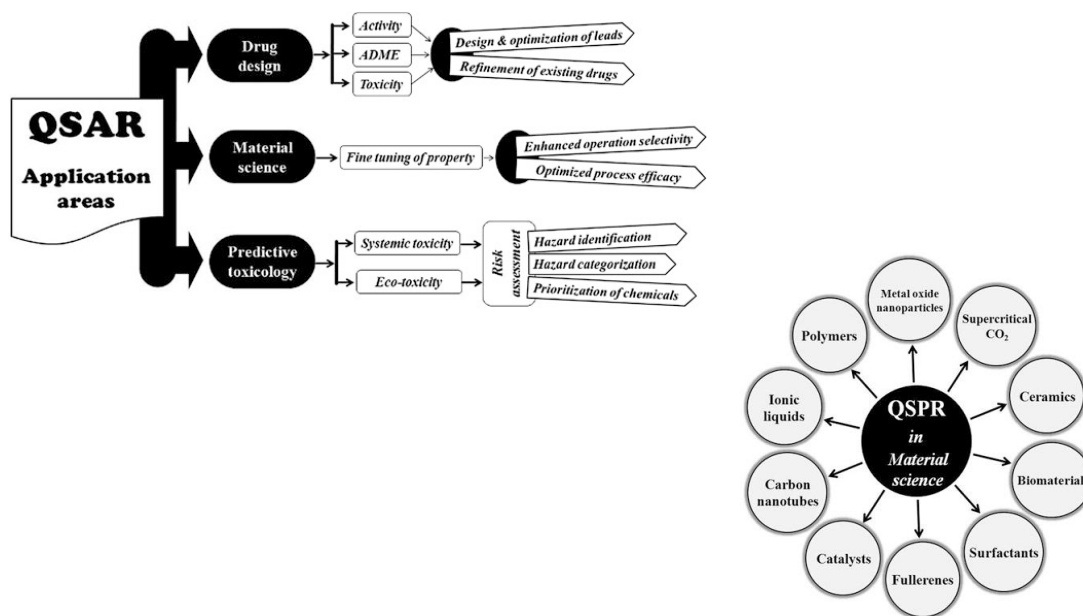


Figure 1: Different application of QSAR/QSPR in various fields

QSPR modeling can be a very good option to predict chemical response using limited resources in any prospective discipline. Hence, we can see that the simple ideology of QSPR, i.e., development of a suitable mathematical correlation between the chemical attributes and a response of interest, can be of significant application to serve the human community. QSAR/QSPR plays an encouraging role in achieving this environmental greenness through the design and development of process-specific chemicals with reduced (or no) hazardous outcomes.

1.2. Nanomaterials

Nanotechnology has achieved tremendous progress in the past several decades. Recently, nanomaterials, which are materials with basic structural units, grains, particles, fibers or other constituent components smaller than 100 nm in at least one dimension [33], have evoked a great amount of attention for improving disease prevention, diagnosis, and treatment. Many authors believe that nanomaterials as materials whose internal structure has nanoscale

dimensions are hardly something new to science. However, it was relatively recently realized that some formations of oxides, metals, ceramics, and other substances are nanomaterials. Nanomaterial is defined as a "*material with any external dimension in the nanoscale or having internal structure or surface structure in the nanoscale*", with nanoscale defined as the "*length range approximately from 1 nm to 100 nm*".

1.2.1 Types of nanomaterials

1.2.1.1 Nanoparticles

a) Fullerenes

- The fullerenes are a class of allotropes of carbon which conceptually are graphene sheets rolled into tubes or spheres. These include the carbon nanotubes (or silicon nanotubes) which are of interest both because of their mechanical strength and also because of their electrical properties.
- For the past decade, the chemical and physical properties of fullerenes have been a hot topic in the field of research and development, and are likely to continue to be for a long time. In April 2003, fullerenes were under study for potential medicinal use: binding specific antibiotics to the structure of resistant bacteria and even target certain types of cancer cells such as melanoma. The October 2005 issue of Chemistry and Biology contains an article describing the use of fullerenes as light-activated antimicrobial agents. In the field of nanotechnology, heat resistance and superconductivity are among the properties attracting intense research.

b) Metal-based nanoparticles

- Inorganic nanomaterials, (e.g. quantum dots, nanowires and nanorods) because of their interesting optical and electrical properties, could be used in optoelectronics [34]. Furthermore, the optical and electronic properties of nanomaterials which depend on their size and shape can be tuned via synthetic techniques. There are the possibilities to use those materials in organic material based optoelectronic devices such as Organic solar cells, OLEDs etc. The operating principles of such devices are governed by photoinduced processes like electron transfer and energy transfer. The

performance of the devices depends on the efficiency of the photoinduced process responsible for their functioning. Therefore, better understanding of those photoinduced processes in organic/inorganic nanomaterial composite systems is necessary in order to use them in optoelectronic devices.

- Nanoparticles or nanocrystals made of metals, semiconductors, or oxides are of particular interest for their mechanical, electrical, magnetic, optical, chemical and other properties [35-36] Nanoparticles have been used as quantum dots and as chemical catalysts such as nanomaterial-based catalysts. Recently, a range of nanoparticles are extensively investigated for biomedical applications including tissue engineering, drug delivery, and biosensor [37-38].

1.2.1.2. One-dimensional nanostructures

The smallest possible crystalline wires with cross-section as small as a single atom can be engineered in cylindrical confinement [39-41]. Carbon nanotubes, a natural semi-1D nanostructure, can be used as a template for synthesis. Confinement provides mechanical stabilization and prevents linear atomic chains from disintegration; other structures of 1D nanowires are predicted to be mechanically stable even upon isolation from the templates [42- 43].

1.2.1.3. Two-dimensional nanostructures

2D materials are crystalline materials consisting of a two-dimensional single layer of atoms. The most important representative graphene was discovered in 2004. Thin films with nanoscale thicknesses are considered nanostructures, but are sometimes not considered nanomaterials because they do not exist separately from the substrate.

1.2.1.4. Bulk nanostructured materials

Some bulk materials contain features on the nanoscale, including nanocomposites, nanocrystalline materials, nanostructured films, and nanotextured surfaces [44].

- Box-shaped graphene (BSG) nanostructure is an example of 3D nonmaterial [45]. BSG nanostructure has appeared after mechanical cleavage of pyrolytic graphite. This nanostructure is a multilayer system of parallel hollow nanochannels located along the

surface and having quadrangular cross-section. The thickness of the channel walls is approximately equal to 1 nm. The typical width of channel facets makes about 25 nm.

1.2.2. Applications of nanomaterials

- Nano materials are used in a variety of manufacturing processes, products and healthcare including paints, filters, lubricant additives and insulation.
- In healthcare Nanozymes are nanomaterials with enzyme-like characteristics [46]. They are an emerging type of artificial enzyme, which have been used for wide applications in such as biosensing, bioimaging, tumor diagnosis [47], antibiofouling and more.
- In paints nanomaterials are used to improve UV protection and improve ease of cleaning [48].
- High quality filters may be produced using nanostructures, these filters are capable of removing particulate as small as a virus as seen in a water filter created by Seldon Technologies. In the air purification field, nano technology was used to combat the spread of MERS in Saudi Arabian hospitals in 2012 [49].
- Nanomaterials are being used in modern and human-safe insulation technologies, in the past they were found in Asbestos-based insulation [50].
- As a lubricant additive, nano materials have the ability to reduce friction in moving parts. Worn and corroded parts can also be repaired with self-assembling anisotropic nanoparticles called TriboTEX [51].
- Nanomaterials can also be used in three-way-catalyst (TWC) applications. TWC converters have the advantage of controlling the emission of nitrogen oxides (NO_x), which are precursors to acid rain and smog [52].
- In core-shell structure, nanomaterials form shell as the catalyst support to protect the noble metals such as palladium and rhodium [52]. The primary function is that the supports can be used for carrying catalysts active components, making them highly dispersed, reducing the use of noble metals, enhancing catalysts activity, and improving the mechanical strength.

1.2.3. Health and safety of nanomaterials

The World Health Organization (WHO) published a guideline on protecting workers from potential risk of manufactured nanomaterials at the end of 2017 [53]. WHO used a precautionary approach as one of its guiding principles, this means that exposure has to be reduced, despite uncertainty about the adverse health effects, when there are reasonable indications to do so. This is highlighted by recent scientific studies that demonstrate a capability of nanoparticles to cross cell barriers and interact with cellular structures [54-55]. In addition, the hierarchy of controls was an important guiding principle. This means that when there is a choice between control measures, those measures that are closer to the root of the problem should always be preferred over measures that put a greater burden on workers, such as the use of personal protective equipment (PPE). WHO commissioned systematic reviews for all important issues to assess the current state of the science and to inform the recommendations according to the process set out in the WHO Handbook for guideline development and the recommendations were rated as "strong" or "conditional" depending on the quality of the scientific evidence, values and preferences, and costs related to the recommendation

In the recent years, QSAR/QSPR modeling have been observed to be useful for modeling response of novel chemicals like ionic liquids, nanoparticles, CNTs etc. thus increasing the area of applications manifold. QSPR modeling has also been found to be beneficial in agricultural sciences, nanotoxicology and in treating environmental pollution. The pollutants discharged into the water bodies from the industries could be modeled against the CNTs (carbon nanotubes) to determine the features which could be essential for uptake by the CNTs. QSPR, modeling of organic pollutants/solvents using adsorption properties/dispersibility index by CNTs can be of great importance for researchers and practitioners. Modeling of property of chemicals encompasses a wide field spanning properties of industrial process chemicals to CNTs. On the other hand, the toxicological modeling is useful for the assessment of all types of chemical toxicity, including that for drugs, pharmaceuticals and nanoparticles. Hence, QSPR modeling has potential contribution to monitor the toxicity of industrial chemicals that are intended to serve various disciplines of biological and material science.

1.3. Role of predictive QSPR models on nanomaterial/nanoparticles.

Performing various experimental procedures using nanoparticles is not only time consuming but also toxic to different organisms. The European Commission (EC) proposed the REACH (Registration, Evaluation and Authorization of Chemicals) system to deal with both existing and new chemical substances. According to it, if testing is to be based on traditional methods, very large numbers of laboratory animals could be needed in response to the REACH system, causing ethical, scientific and logistical problems that would be incompatible with the time-schedule envisaged for testing. EC tried to minimize testing on animals but failed. The non-animal methods i.e., QSARs could be used in a tiered approach to provide a rapid and scientifically justified basis for the risk assessment of chemicals for their toxic effects in humans and identification of various properties of nanoparticles. The information obtained from Quantitative structure-activity relationships (QSARs) can be applied in substance-tailored testing schemes which can evaluate the reproductive toxicity of chemicals that can have a stronger impact on the total number of animals bred for testing under REACH.

Recent studies have reported predictive QSPR models on various properties and biological/toxicological responses of nanoparticles. Modeling based on properties enables the design and development of purpose specific efficient analogues, while models on toxicity response allow the user to capture specific information on the hazardous attribute. However, considering the scope of this dissertation, we would like to present an account on some of the representative published QSPR models on toxicity and property of nanoparticles.

Apul et al [56] performed Quantitative Structure–Activity Relationship (QSAR) and Linear Solvation Energy Relationship (LSER) techniques to develop predictive models for adsorption of organic contaminants by multi-walled carbon nanotubes (MWCNTs). They worked with 46 aromatic compounds and the generated QSAR ($r^2 = 0.88$), and LSER ($r^2 = 0.83$) equations were validated externally using an independent validation data set of 30 aromatic compounds. External validation accuracies indicated the success of parameter selection, data fitting ability, and the prediction strength of the developed models. Finally, the combination of training and validation data were used to obtain a combined LSER equation ($r^2 = 0.83$) that would be used for predicting adsorption of a wide range of low molecular weight aromatics by MWCNTs.

Wang et al. [57] developed a 3D-QSAR model for the adsorption coefficient of 39 aromatic compounds on multi-walled carbon nanotubes to have an understanding of the relationship between adsorption coefficients and physicochemical properties of aromatic compounds. So, a 3D QSAR (quantitative structure–property relationship) model was developed by the utilization of 3D molecular structures of 39 aromatic compounds. In the model development process, three different learning approaches, multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM), were used. The validation results showed that SVM- and ANN-based models resulted in a better agreement between predicted and measured values, with the coefficient of determination (R^2) of 0.8317 and 0.7829, than the MLR-based model with R^2 of 0.5093.

Nano-QTTR model with metal oxide nanoparticles was developed by Kar et al. [58] for determining the toxicity of nanoparticles towards *Escherichia coli* and human keratinocyte cell line. The developed model proved that nano-QTTR/QSPR can be employed to assess the discriminatory features for cytotoxicity of metal oxide nanoparticles. Informative illustrations of the contributing mechanisms of toxic action of the metal oxide nanoparticles to the HaCaT cell line as well as to the *E. coli* are identified from the developed nano quantitative toxicity–toxicity relationship (nano-QTTR) models.

In another paper, De et al. [59] performed quantitative structure–toxicity relationship (QSTR) approach on metal oxides to have a better understanding of the toxicities of metal oxide nanoparticles with different species (*E. coli*, a human keratinocyte cell line (HaCaT) and zebrafish embryos) along with the identification of the major mechanism(s) for such toxicities. The authors employed the developed 1st and 2nd generation periodic table-based descriptors. These models further helped in extrapolating toxicity when the data for one species are available and the data for other species are unavailable.

Puzyn et al. [60] performed quantitative structure–activity relationship (QSAR) method to predict the toxicity of various metal oxides. They developed a QSPR model to describe the cytotoxicity of 17 different types of metal oxide nanoparticles to bacteria *Escherichia coli*. The model reliably predicted the toxicity of all considered compounds, and the methodology is expected to provide guidance for the future design of safe nanomaterials.

Gajewicz et al. [61] developed nano-QSAR model, which provided governance over the toxicity of metal oxide nanoparticles to the human keratinocyte cell line (HaCaT). The combined experimental–theoretical studies allowed the development of an interpretative nano-QSAR model describing the toxicity of 18 nano-metal oxides to the HaCaT cell line, which is a common in vitro model for keratinocyte response during toxic dermal exposure. The comparison of the toxicity of metal oxide nanoparticles to bacteria *Escherichia coli* (prokaryotic system) and a human keratinocyte cell line (eukaryotic system), resulted in the hypothesis that different modes of toxic action occur between prokaryotic and eukaryotic systems

Although nanotechnology has given solutions to various problems there are still many gaps in available experimental data devoted to property of nanoparticles (engineered) that are already available in the market. Therefore, there is a need to reduce the cost and time by bridging the existing data gaps by applying combined experimental and computational approaches. So, with the rapid increase of nanomaterials it has become essential to determine the property as well as the toxicity caused by it to different organisms. Here, we have developed QSPR models in accordance to the OECD guideline and therefore predicting the possible mechanism of toxicity of the nanoparticles or nanomaterials in order to minimize the toxic effect and enhance the property of the nanoparticles so that their property could be used widely in various fields with minimum toxicity.

CHAPTER - 2

Present work

2. PRESENT WORK

Nanotechnology has taken the frontline in the modern world of science [62]. Nanoscale materials are substances comprising one or more features less than 100 nm in at least one dimension. In theory, nanoparticle can be engineered from any substance like semiconductors nanocrystals, organic dendrimers and carbon fullerenes and possess properties like electrical, thermal, mechanical which are highly desirable in commercial, medical and environmental sectors [63]. Sensitivity enhancement and miniaturization is favored due to its large surface to volume ratio. NPs possess unique optical, electronic and magnetic properties, depending on their core materials. Nanoparticles (NPs), which have unique chemical and physical properties, are promising materials in our overall strategies to detect and remediate environmental pollutants [64]. These properties changes with their surrounding chemical environment providing a foundation for pollutant sensing. To further improve their sensitivity, NPs can also be incorporated with a wide range of small organic molecules or polymers by surface modifications. Wastewater from many industries such as metallurgical, tannery, chemical manufacturing, mining, battery manufacturing industries, etc. contains one or more toxic metal ions and organic chemicals (OCs) which could be removed by nanotechnology. Nanoparticles use of may endanger human health through the potential induction of cytogenetic, mutagenic, or neurotoxic health effects [64-66]. It is necessary to remove these nanoparticles and organic chemicals from the wastewaters and terrestrial surface before releasing into the environment, because there is possibility of entry of toxic metal ions and OCs into food chain through waste discharges into water bodies and terrestrial. As the conventional process or method of risk assessment and removal of pollutants from water is expensive and time consuming, computational based assessment method should be alternatively used which not only reduces the number of experiments and also the cost of consumable reagents.

Development of predictive models in the form of QSPR/QSTR analysis provides a well validated rational platform for the determination of activity/toxicity of all these new chemicals and to fill data gaps of old chemicals as well.

In this work, we have performed QSPR models of adsorption property of diverse organic pollutants with Dragon and PaDel descriptors and determined their key aspects which influenced their adsorption by the CNTs and investigated their contribution to the model using approach of *in silico* method. The data set taken in account was adsorption coefficient (K_{∞}) which was directed to multi walled nanotubes (MWNTs) of OCs pollutants by CNTs. Here, we have correlated the adsorption coefficient (K_{∞}) with the molecular descriptors that encode the various aspects of molecular structure.

2.1. Study 1: Dataset 1

The pollution due to the indiscriminate disposal of different wastes has been causing worldwide concern. Wastewater from many industries such as metallurgical, tannery, chemical manufacturing, mining, battery manufacturing industries, etc. contains one or more toxic metal ions and organic chemicals (OCs), which need to be removed for making pollution free water environment. MWCNTS (multi-walled nanotubes) can be used to fulfill such purpose.

The reason for selecting CNTs is because of its large specific surface area, light mass, hollow and porous structure of CNTs which makes it applicable for removal of hazardous pollutant from both aqueous solutions and gas stream [65-68] CNTs can absorb pollutants through various mechanism such as van der Waals forces, electrostatic interactions, interactions π - π interaction, hydrogen bonding and electro phobic interaction, the mechanism depends on the type of organic compounds whether polar or non polar [69]. Contributions to adsorption are also made by π -electron polarizability.

The data set consisting of adsorption affinity properties (k_{∞}) of 59 organic contaminants by multi-walled carbon nanotubes (MWCNTs) were modeled to understand the essential features responsible for the adsorptive activity and the requisite features of molecules that is important to increase or decrease the adsorption coefficient of the organic pollutants. The QSPR models would provide an important guidance for the chemists to predict adsorption of organic pollutants by multi walled carbon nanotubes (MWNTs) theoretically thereby saving the time and resources involved in the experimental determination.

Table 1: Compounds name with respective experimental log k_{∞} values

No.	Chemical Name	Expt. (log K_{∞})	No.	Chemical Name	Expt. (log K_{∞})
1	pyrene	4.01	31	Bromobenzene	0.5
2*	naphthalene	1.63	32	propylbenzene	0.76
3*	1-naphthol	0.76	33	4-chlorotoluene	0.82
4	Biphenyl	2.05	34	Benzonitrile	0.04
5	2-phenylphenol	1.63	35*	4-fluorophenol	-0.32
6	benzene	-0.45	36	benzyl alcohol	-0.9
7*	chlorobenzene	-0.33	37	iodobenzene	0.88
8	1,2,4-trichlorobenzene	1.17	38	Acetophenone	0.26
9	nitrobenzene	0.33	39	3-methylphenol	0.08
10	2,4-dinitrotoluene	2.38	40*	methyl benzoate	0.7
11	phenol	-0.54	41*	4-chloroanisole	1.07
12	catechol	0.21	42	phenethyl alcohol	-0.46
13	pyrogallol	1.18	43	3-methylbenzyl alcohol	-0.15
14	2,4,6-trichlorophenol	1.43	44	4-ethylphenol	0.62
15	3-nitrotoluene	1.03	45	3,5-dimethylphenol	0.49
16	4-nitrophenol	0.77	46	ethyl benzoate	1.14
17*	aniline	-0.77	47	methyl 2-methylbenzoate	1.12
18	4-chloroaniline	-0.66	48	3-chlorophenol	0.62
19	2-nitroaniline	1.6	49*	4-nitrotoluene	1.44
20*	3-nitroaniline	0.72	50	4-chloroacetophenone	1.28
21*	4-nitroaniline	0.95	51	3-bromophenol	0.79
22	4-methylphenol	0.06	52	1-methylnaphtalene	1.89
23	2-chlorophenol	0.08	53	2-dichlorobenzene	0.56
24	4-chlorophenol	0.74	54	3-dichlorobenzene	0.65
25	2,4-dichlorophenol	0.96	55*	4-dichlorobenzene	0.51
26	2-nitrophenol	0.56	56*	isophorone	0.01
27	3-nitrophenol	0.92	57	2-chloronaphthalene	2.73

28	1,3-dinitrobenzene	1.46	58	azobenzene	2.72
29	ethylbenzene	0.19	59*	Phenanthrene	3.29
30*	4-xylene	0.26			

*test set compounds

Models were developed using Dragon and PaDel descriptors. Data set division, method like *K-Medoid* [70] was applied for dataset division into training (model development) and test set (prediction). The models were developed using methods such as stepwise-Multiple Linear Regression (MLR) [71] method in MINITAB software [72] and Best sub set selection software. Finally, the selected models were run using intelligent consensus predictor (ICP) tool developed in our laboratory [73] to explore whether the quality of predictions of external compounds can be enhanced through an “intelligent” selection of multiple models (in this paper selected five models). The other validation parameters were also checked.

2.2. Study 2: Dataset 2

Metal oxide nanoparticles (NPs) have unique property due to their size and high density of edge or corner sites. There is an increase use of nanoparticles in different areas like space technology, pharmacy, environmental engineering, cosmetic, stain-resistant clothing, environmental monitoring and so on [74-75]. By the end of 2019, its worldwide market is estimated to be \$79.8 billion [76]. It has been noted that metal oxide induces toxicity to some organisms [77] and is believed that exponential growth in use of nanoparticle may endanger human health through the potential induction of cytogenetic, mutagenic, or neurotoxic health effects [78-79]. The proposed procedure (method) should be developed which not only should be able to perform risk assessment but also to reduce the extensive animal testing and provide detail information about toxicity mechanism at molecular level.

Usually, Quantitative structure–activity relationship (QSAR) is not applied widely yet for predicting the toxicity or physiological activity of the heterogeneous NPs. Due to lack of appropriate nano-descriptors which could explain the characteristics of the NPs, it is very difficult to develop nano-QSPR or nano-QSAR models to determine the cytotoxicity for the heterogeneous NPs. . Thus, we have employed a novel approach for development of nano-

QSTR models to determine the cytotoxicity for the heterogeneous NPs using simple periodic table additive descriptors for mixture compounds. Based on the additive descriptors, we have developed mono parametric based QSTR models of 34 TiO₂-based NPs modified with (poly) metallic clusters of novel metals (Au, Ag, Pt) to assess the cytotoxicity (log EC₅₀) towards Chinese Hamster Ovary cell line.

The model was developed using period table based descriptors. After dataset division with *K-Medoid* clustering technique, Best subset selection using mixed descriptors (33 descriptors) was performed and five mono parametric LR models selected with single descriptor based on the MAE based criteria then PLS regression(with one latent variable) was performed for the final model selection. The results were statistically good and aimed at providing statistically robust predictions for the toxicological activity of the compounds. This model can be used as an efficient tool to assess the toxicity with physiological property of the new heterogeneous NPs in future.

CHAPTER - 3

Materials and methods

3. MATERIALS AND METHODS

The present dissertation aims in implementing a transparent methodological framework for the development of predictive QSAR models using simple descriptors. We have endeavored to maintain explicitness for computation of the descriptors, thinning of the variable matrix, selection of potential features as well as judgment of robustness and predictivity of the models. Here, we have presented the details of the dataset along with their activity and toxicity data which is being employed for in silico studies i.e. QSPR. The done was divided into the following parts as follows

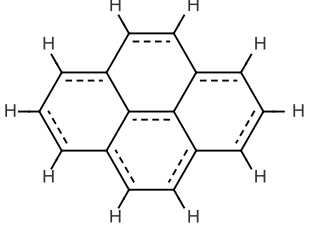
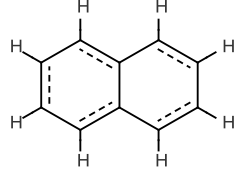
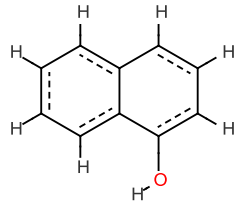
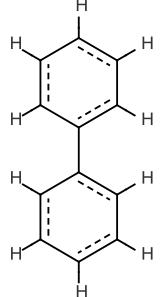
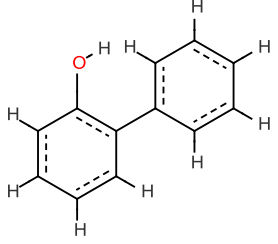
- Details of datasets consisting chemical structures along with their activity or toxicity data.
- General description of methods implemented for developing QSAR models
- Study wise specific description of methodologies utilized in each study

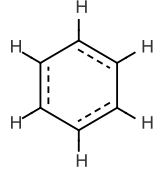
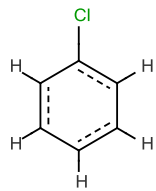
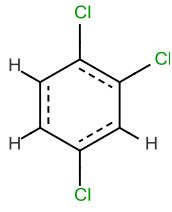
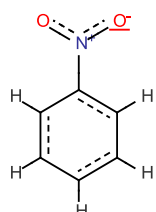
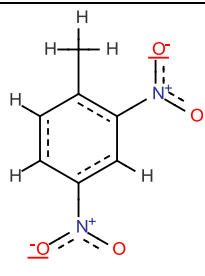
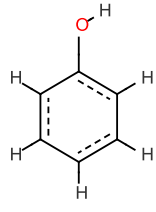
3.1. Details of datasets consisting chemical structures along with their activity or toxicity data

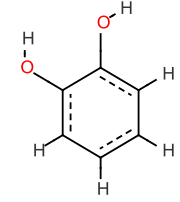
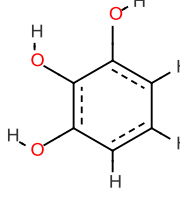
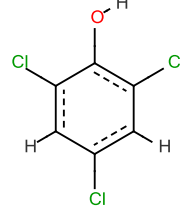
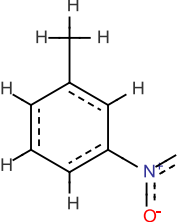
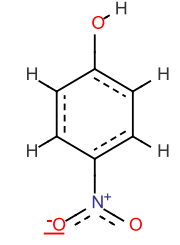
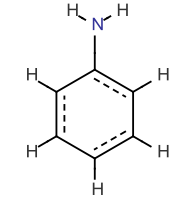
3.1.1. Dataset I (study 1)

The dataset defined adsorption affinity properties (k_{∞}) of 59 organic contaminants by multi-walled carbon nanotubes (MWCNTs). All the endpoint values were taken in logarithmic scale for the modeling purposes. The data sets mainly contain adsorption data for synthetic organic compounds like pyrene, naphthalene, phenol, benzene, aniline, benzoate, chloroanisole, alcohol, acetophenone, isophoron, phenanthrene dicamba, atrazine, carbamazepine, pyrimidinone, acetamide, piperidine, propionitrile, acrylic acid, thiodiethanol, ethanol amine, cyclopentanone, acetone and ethylene glycol derivatives. K_{∞} is adsorption coefficients which could be obtained from isotherm data. K_{∞} is the ratio of q_e and C_e (q_e and C_e are solid and liquid phase equilibrium concentration respectively, at infinite dilution conditions with an average of 0.2% aqueous solubility). The data set details are given in **Table 3.1**.

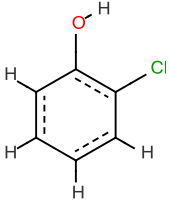
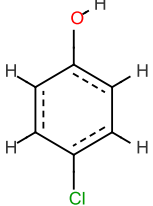
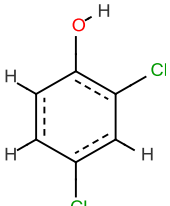
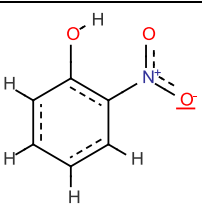
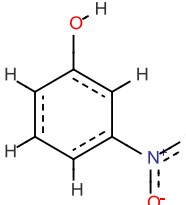
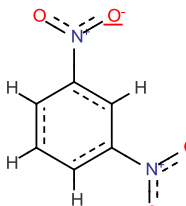
Table 3.1: Compounds name with respective experimental $\log K_{\infty}$ values

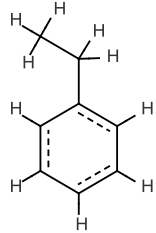
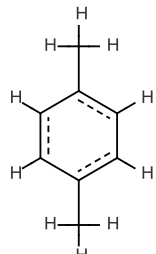
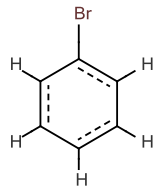
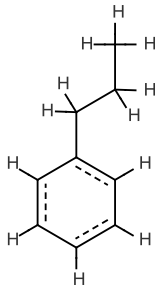
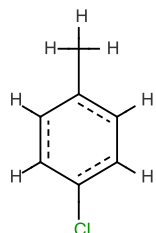
Sl. no	Chemical Name	Chemical Structure	Expt. $\log K_{\infty}$
1	pyrene		4.01
2*	naphthalene		1.63
3*	1-naphthol		0.76
4	biphenyl		2.05
5	2-phenylphenol		1.63

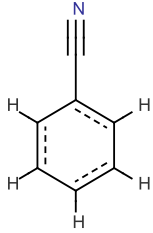
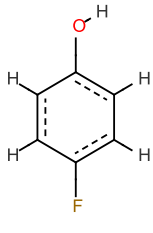
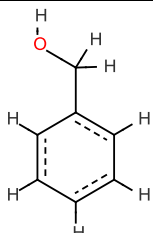
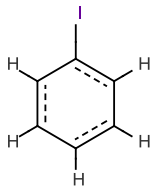
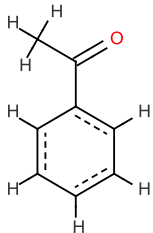
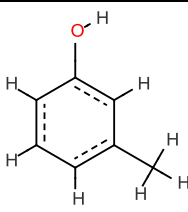
6	benzene		-0.45
7*	chlorobenzene		-0.33
8*	1,2,4-trichlorobenzene		1.17
9	nitrobenzene		0.33
10	2,4-dinitrotoluene		2.38
11	phenol		-0.54

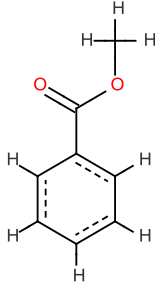
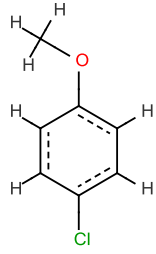
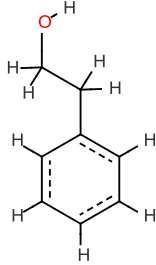
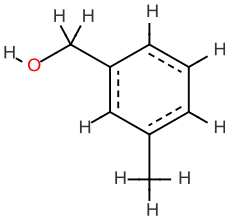
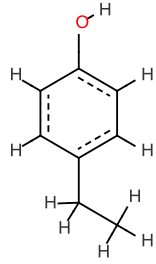
12	catechol		0.21
13	pyrogallol		1.18
14	2,4,6-trichlorophenol		1.43
15	3-nitrotoluene		1.03
16	4-nitrophenol		0.77
17*	aniline		-0.77

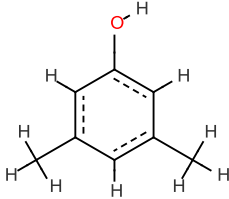
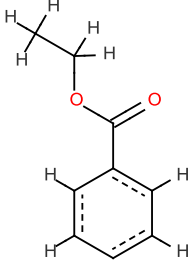
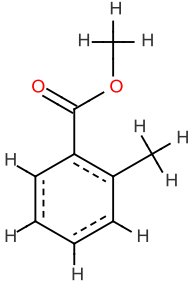
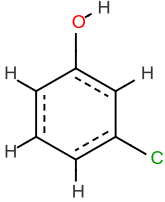
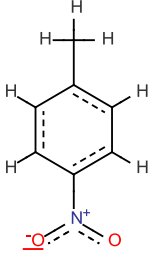
18	4-chloroaniline		-0.66
19	2-nitroaniline		1.6
20*	3-nitroaniline		0.72
21*	4-nitroaniline		0.95
22	4-methylphenol		0.06

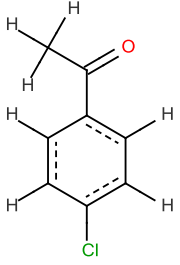
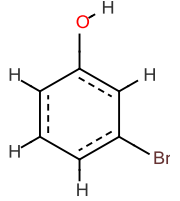
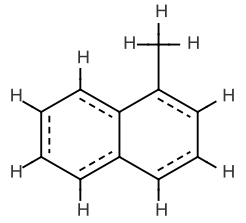
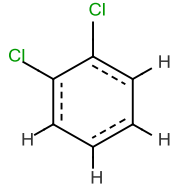
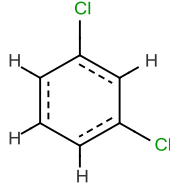
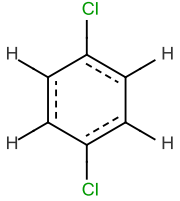
23	2-chlorophenol		0.08
24	4-chlorophenol		0.74
25	2,4-dichlorophenol		0.96
26	2-nitrophenol		0.56
27	3-nitrophenol		0.92
28	1,3-dinitrobenzene		1.46

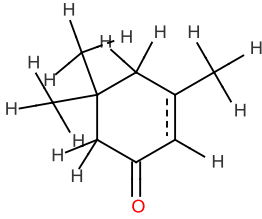
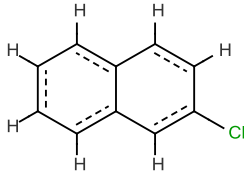
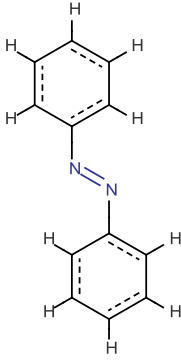
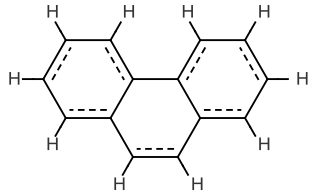
29	ethylbenzene		0.19
30*	4-xylene		0.26
31	bromobenzene		0.5
32	propylbenzene		0.76
33	4-chlorotoluene		0.82

34	benzonitrile		0.04
35*	4-fluorophenol		-0.32
36	benzyl alcohol		-0.9
37	iodobenzene		0.88
38	acetophenone		0.26
39	3-methylphenol		0.08

40*	methyl benzoate		0.7
41*	4-chloroanisole		1.07
42	phenethyl alcohol		-0.46
43	3-methylbenzyl alcohol		-0.15
44	4-ethylphenol		0.62

45	3,5-dimethylphenol		0.49
46	ethyl benzoate		1.14
47	methyl 2-methylbenzoate		1.12
48	3-chlorophenol		0.62
49*	4-nitrotoluene		1.44

50	4-chloroacetophenone		1.28
51	3-bromophenol		0.79
52	1-methylnaphtalene		1.89
53	2-dichlorobenzene		0.56
54	3-dichlorobenzene		0.65
55*	4-dichlorobenzene		0.51

56*	isophorone		0.01
57	2-chloronaphthalene		2.73
58	azobenzene		2.72
59*	phenanthrene		3.29

*test set compounds

3.1.2. Dataset II (study 2)

We have developed QSTR models of 34 TiO₂ [80] NPs (nanoparticles) modified with varying amount and types of mixture of noble metals like Ag, Au and Pt (expressed in mole %). The cytotoxicity data towards the Chinese hamster ovary cell line (CHO-K1, ATCC® CCL-61™) was expressed in $-\log EC_{50}$ (negative logarithm of EC₅₀) for development of QSTR models as given in **Table 3.2**. For the purpose of modeling, all the nanoparticles were

utilized and no single NPs were omitted. All the NPs used in the QSTR modeling were obtained from micro emulsion method [80].

Table 3.2: The details of the dataset with observed cytotoxicity values

No.	Element	Toxicity (-log EC ₅₀)
1*	0.1Pt	4.53
2	0.1Au	4.56
3	0.25Au	4.62
4	0.5Ag_0.1Pt	4.64
5*	0.25Au_0.25Pt	4.66
6	0.05Au_0.05Pt	4.67
7	0.25Pt	4.67
8	0.5Au_0.5Pt	4.68
9*	0.1Au_0.1Pt	4.68
10	0.5Au_0.25Pt	4.7
11	0.1Au_0.25Pt	4.7
12	1.25Pt	4.71
13*	0.5Ag	4.72
14*	0.5Pt	4.73
15	0.5Ag_0.25Pt	4.73
16	0.5Au_0.1Pt_400	4.75

17	0.25Au_0.1Pt	4.76
18	1.5Ag_0.1Pt	4.84
19	1.5Ag	4.89
20*	0.5Ag_0.5Pt	4.94
21	1.5Ag_0.25Pt	5.01
22*	2.5Ag_0.1Pt	5.06
23	1.5Ag_0.5Pt	5.26
24	2.5Ag_0.5Pt	5.32
25	2.5Ag	5.35
26	2.5Ag_0.25Pt	5.37
27*	4.5Ag_0.1Pt	5.54
28	6.5Ag_0.1Pt	5.63
29	4.5Ag_0.25Pt	5.65
30	4.5Ag_0.5Pt	5.65
31*	4.5Ag	5.7
32	6.5Ag_0.5Pt	5.8
33	6.5Ag_0.25Pt	5.84
34	6.5Ag	5.88

* test set compounds

3.2. General description of methods implemented for developing QSAR models

3.2.1. Descriptor calculation

Descriptors are the ultimate mathematical representation of the molecules extracting essential chemical features required for the exertion of a particular response of interest by a chemical compound. According to Todeschini and Consonni [9] a chemical descriptor is defined as "The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment". Descriptors are the numerical representation which is used to correlate the biological and physiological activity with its molecular structure. QSPR modeling helps in encoding of the chemical compounds to the vector numerical descriptors.

In our work, we have performed the calculation 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, atom type E-states, 2D atom pairs, molecular properties (using dragon software version 6)[81] and ETA indices (using PaDEL descriptor software) [82]. Our models showed satisfactory results with 2D descriptors only and the descriptors used. We have also calculated Integral additive descriptors which are a new approach by Mikolajczyk et.al (2018) [80].

Table 3.3: Descriptors for QSPR studies from Dragon

Sl. No	Category of Descriptors	Notation of descriptors
1	Constitutional indices	MW, AMW, Sv, Se, Sp, Si, Mv, Me, Mp, Mi, nAT, nSK, nBT, nBO, nBM, SCBO, RBN, RBF, nDB, nTB, nAB, nH, nC, nN, nO, nP, nS, nF, nCL, nBR, nI, nB, nHM, nHet, nX, H%, C%, N%, O%, X%, nCsp3, nCsp2, nCsp
2	Atom-type E-state indices	SsCH3, SdCH2, SssCH2, StCH, SdsCH, SaaCH, SsssCH, SddC, StsC, SdssC, SaasC, SaaaC, SssssC, SsNH2, SssNH, SdNH, SssN, SdsN, SaaN, StN, SsNH3+, SssNH2+, SdNH2+, SssNH+, SssssN+, SddsN, SaasN, SaaNH, SsOH, SdO, SssO, SaaO, SsPH2, SssPH, SsssP, SdsssP, SddsP, SssssP, SsSH, SdS, SssS, SaaS, SdssS, SddssS, SsssssS, SsF, SsCl, SsBr, SsI, SsLi, SsBH2, SssBH, SsssB, SssssB-, NsCH3, NdCH2, NssCH2, NtCH,

		NdsCH, NaaCH, NsssCH, NddC, NtsC, NdssC, NaasC, NaaaC, NssssC, NsNH2, NssNH, NdNH, NssN, NdsN, NaaN, NtN, NsNH3+, NssNH2+, NdNH2+, NssNH+, NssssN+, NddsN, NaasN, NaaNH, NsOH, NdO, NssO, NaaO, NsPH2, NssPH, NsssP, NdsssP, NddsP, NsssssP, NsSH, NdS, NssS, NaaS, NdssS, NddssS, NsssssS, NsF, NsCl, NsBr, NsI, NsLi, NsBH2, NssBH, NsssB, NssssB
3	Functional group counts	nCp, nCs, nCt, nCq, nCrS, nCrt, nCrq, nCar, nCbH, nCb-, nCconj, nR=Cp, nR=Cs, nR=Ct, n=C=, nR#CH/X, nR#C-, nROCN, nArOCN, nRNCO, nArNCO, nRSCN, nArSCN, nRNCS, nArNCS, nRCOOH, nArCOOH, nRCOOR, nArCOOR, nRCONH2, nArCONH2, nRCONHR, nArCONHR, nRCONR2, nArCONR2, nROCON, nArOCON, nRCOX, nArCOX, nRCSOH, nArCSOH, nRCSSH, nArCSSH, nRCOSR, nArCOSR, nRCSR, nArCSSR, nRCHO, nArCHO, nRCO, nArCO, nCONN, nC=O(O)2, nN=C-N<, nC(=N)N2, nRC=N, nArC=N, nRCNO, nArCNO, nRNH2, nArNH2, nRNHR, nArNHR, nRNR2, nArNR2, nN-N, nN=N, nRCN, nArCN, nN+, nNq, nRNHO, nArNHO, nRNNOX, nArNNOX, nRNO, nArNO, nRNO2, nArNO2, nN(CO)2, nC=N-N<, nROH, nArOH, nOHp, nOHs, nOHT, nROR, nArOR, nROX, nArOX, nO(C=O)2, nH2O, nSH, nC=S, nRSR, nRSSR, nSO, nS(=O)2, nSOH, nSOOH, nSO2OH, nSO3OH, nSO2, nSO3, nSO4, nSO2N, nPO3, nPO4, nPR3, nP(=O)O2R, nP(=O)R3/nPR5, nCH2RX, nCHR2X, nCR3X, nR=CHX, nR=CRX, nR#CX, nCHRX2, nCR2X2, nR=CX2, nCRX3, nArX, nCXr, nCXr-, nCconjX, nAziridines, nOxiranes, nThiranes, nAzetidines, nOxetanes, nThioethanes, nBeta-Lactams, nPyrrolidines, nOxolanes, nH-Thiophenes, nPyrroles, nPyrazoles, nImidazoles, nFuranes, nThiophenes, nOxazoles, nIsoxazoles, nThiazoles, nIsothiazoles, nTriazoles, nPyridines, nPyridazines, nPyrimidines, nPyrazines, n135-Triazines, n124-Triazines, nHDon, nHAcc, nHBonds
4	2D Atom Pairs	T(N..N), T(N..O), T(N..S), T(N..P), T(N..F), T(N..Cl), T(N..Br), T(N..I), T(O..O), T(O..S), T(O..P), T(O..F), T(O..Cl), T(O..Br), T(O..I), T(S..S), T(S..P), T(S..F), T(S..Cl), T(S..Br), T(S..I), T(P..P), T(P..F), T(P..Cl), T(P..Br), T(P..I), T(F..F), T(F..Cl), T(F..Br), T(F..I), T(Cl..Cl), T(Cl..Br), T(Cl..I), T(Br..Br), T(Br..I), T(I..I), B01[C-C], B01[C-N], B01[C-O], B01[C-S], B01[C-P], B01[C-F], B01[C-Cl], B01[C-Br], B01[C-I], B01[C-B], B01[C-Si], B01[C-X], B01[N-N], B01[N-O], B01[N-S], B01[N-P], B01[N-F], B01[N-Cl], B01[N-Br], B01[N-I], B01[N-B], B01[N-Si], B01[N-X], B01[O-O], B01[O-S], B01[O-P], B01[O-F], B01[O-Cl], B01[O-Br], B01[O-I], B01[O-B], B01[O-Si], B01[O-X], B01[S-S], B01[S-P], B01[S-F], B01[S-Cl], B01[S-Br], B01[S-I], B01[S B], B01[S-Si], B01[S-X], B01[P-P], B01[P-F], B01[P-Cl], B01[P-Br], B01[P-I], B01[P-B], B01[P-Si], B01[P-X], B01[F-F], B01[F-Cl], B01[F-Br], B01[F-I], B01[F-B], B01[F-Si], B01[F-X], B01[Cl- Cl], B01[Cl-Br], B01[Cl-I], B01[Cl-B], B01[Cl-Si], B01[Cl-X], B01[Br-Br], B01[Br-I], B01[Br-B], B01[Br-Si], B01[Br-X], B01[I-I], B01[IB], B01[I-Si], B01[I-X], B01[B-B], B01[B-Si], B01[B-X], B01[Si-Si], B01[Si-X], B01[X-X], B02[C-C], B02[C-N], B02[C-O], B02[C-S], B02[C-P], B02[C-F], B02[C-Cl], B02[C-Br], B02[C-I], B02[C-B], B02[C-Si],

		<p>B02[C-X], B02[N-N], B02[N-O], B02[N-S], B02[N-P], B02[N-F], B02[N-Cl], B02[N-Br], B02[N-I], B02[N-B], B02[N-Si], B02[N-X], B02[O-O], B02[O-S], B02[O-P], B02[O-F], B02[O-Cl], B02[O-Br], B02[O-I], B02[O-B], B02[O-Si], B02[O-X], B02[S-S], B02[S-P], B02[S-F], B02[S-Cl], B02[S-Br], B02[S-I], B02[S-B], B02[S-Si], B02[S-X], B02[P-P], B02[P-F], B02[P-Cl], B02[P-Br], B02[P-I], B02[P-B], B02[P-Si], B02[P-X], B02[F-F], B02[F-Cl], B02[F-Br], B02[F-I], B02[F-B], B02[F-Si], B02[F-X], B02[Cl-Cl], B02[Cl-Br], B02[Cl-I], B02[Cl-B], B02[Cl-Si], B02[Cl-X], B02[Br-Br], B02[Br-I], B02[Br-B], B02[Br-Si], B02[Br-X], B02[I-I], B02[I-B], B02[I-Si], B02[I-X], B02[B-B], B02[B-Si], B02[B-X], B02[Si-Si], B02[Si-X], B02[X-X], B03[C-C], B03[C-N], B03[C-O], B03[C-S], B03[C-P], B03[C-F], B03[C-Cl], B03[C-Br], B03[C-I], B03[C-B], B03[C-Si], B03[CX], B03[N-N], B03[N-O], B03[N-S], B03[N-P], B03[N-F], B03[N-Cl], B03[N-Br], B03[N-I], B03[N-B], B03[N-Si], B03[N-X], B03[O-O], B03[O-S], B03[O-P], B03[O-F], B03[O-Cl], B03[O-Br], B03[O-I], B03[O-B], B03[O-Si], B03[O-X], B03[S-S], B03[S-P], B03[S-F], B03[S-Cl], B03[S-Br], B03[S-I], B03[S-B], B03[S-Si], B03[S-X], B03[P-P], B03[P-F], B03[P-Cl], B03[P-Br], B03[P-I], B03[P-B], B03[P-Si], B03[P-X], B03[F-F], B03[F-Cl], B03[F-Br], B03[F-I], B03[F-B], B03[F-Si], B03[F-X], B03[Cl-Cl], B03[Cl-Br], B03[Cl-I], B03[Cl-B], B03[Cl-Si], B03[Cl-X], B03[Br-Br], B03[Br-I], B03[Br-B], B03[Br-Si], B03[Br-X], B03[I-I], B03[I-B], B03[I-Si], B03[I-X], B03[B-B], B03[B-Si], B03[B-X], B03[Si-Si], B03[Si-X], B03[X-X], B04[C-C], B04[C-N], B04[C-O], B04[C-S], B04[C-P], B04[C-F], B04[C-Cl], B04[C-Br], B04[C-I], B04[C-B], B04[C-Si], B04[C-X], B04[N-N], B04[N-O], B04[N-S], B04[N-P], B04[N-F], B04[N-Cl], B04[N-Br], B04[N-I], B04[N-B], B04[N-Si], B04[N-X], B04[O-O], B04[O-S], B04[O-P], B04[O-F], B04[O-Cl], B04[O-Br], B04[O-I], B04[O-B], B04[O-Si], B04[O-X], B04[S-S], B04[S-P], B04[S-F], B04[S-Cl], B04[S-Br], B04[S-I], B04[S-B], B04[S-Si], B04[S-X], B04[P-P], B04[P-F], B04[P-Cl], B04[P-Br], B04[P-I], B04[P-B], B04[P-Si], B04[P-X], B04[F-F], B04[F-Cl], B04[F-Br], B04[F-I], B04[F-B], B04[F-Si], B04[FX], B04[Cl-Cl], B04[Cl-Br], B04[Cl-I], B04[Cl-B], B04[Cl-Si], B04[Cl-X], B04[Br-Br], B04[Br-I], B04[Br-B], B04[Br-Si], B04[Br-X], B04[I-I], B04[I-B], B04[I-Si], B04[I-X], B04[B-B], B04[B-Si], B04[B-X], B04[Si-Si], B04[Si-X], B04[X-X], B05[C-C], B05[C-N], B05[CO], B05[C-S], B05[C-P], B05[C-F], B05[C-Cl], B05[C-Br], B05[C-I], B05[C-B], B05[C-Si], B05[C-X], B05[N-N], B05[N-O], B05[N-S], B05[N-P], B05[N-F], B05[N-Cl], B05[N-Br], B05[N-I], B05[N-B], B05[N-Si], B05[N-X], B05[O-O], B05[O-S], B05[O-P], B05[O-F], B05[O-Cl], B05[O-Br], B05[O-I], B05[O-B], B05[O-Si], B05[O-X], B05[S-S], B05[S-P], B05[S-F], B05[S-Cl], B05[S-Br], B05[S-I], B05[SB], B05[S-Si], B05[S-X], B05[P-P], B05[P-F], B05[P-Cl], B05[P-Br], B05[P-I], B05[P-B], B05[P-Si], B05[P-X], B05[F-F], B05[F-Cl], B05[F-Br], B05[F-I], B05[F-B], B05[F-Si], B05[F-X], B05[Cl-Cl], B05[Cl-Br], B05[Cl-I], B05[Cl-B], B05[Cl-Si], B05[Cl-X], B05[Br-Br], B05[Br-I], B05[Br-B], B05[Br-Si], B05[Br-X], B05[I-I], B05[I-B], B05[I-Si], B05[I-X], B05[B-B], B05[B-Si], B05[B-X], B05[Si-Si], B05[Si-X], B05[X-X], B06[C-C], B06[C-N], B06[C-O], B06[C-</p>
--	--	--

		<p>S], B06[C-P], B06[C-F], B06[C-Cl], B06[C-Br], B06[C-I], B06[C-B], B06[C-Si], B06[C-X], B06[N-N], B06[N-O], B06[N-S], B06[N-P], B06[N-F], B06[N-Cl], B06[N-Br], B06[N-I], B06[N-B], B06[N-Si], B06[N-X], B06[O-O], B06[O-S], B06[O-P], B06[O-F], B06[O-Cl], B06[O-Br], B06[O-I], B06[O-B], B06[O-Si], B06[O-X], B06[S-S], B06[S-P], B06[S-F], B06[S-Cl], B06[S-Br], B06[S-I], B06[S-B], B06[S-Si], B06[S-X], B06[P-P], B06[P-F], B06[P-Cl], B06[P-Br], B06[P-I], B06[P-B], B06[P-Si], B06[P-X], B06[F-F], B06[F-Cl], B06[F-Br], B06[F-I], B06[F-B], B06[F-Si], B06[F-X], B06[Cl-Cl], B06[Cl-Br], B06[Cl-I], B06[Cl-B], B06[Cl-Si], B06[Cl-X], B06[Br-Br], B06[Br-I], B06[Br-B], B06[Br-Si], B06[Br-X], B06[I-I], B06[I-B], B06[I-Si], B06[I-X], B06[B-B], B06[B-Si], B06[B-X], B06[Si-Si], B06[Si-X], B06[X-X], B07[C-C], B07[C-N], B07[C-O], B07[C-S], B07[C-P], B07[C-F], B07[C-Cl], B07[C-Br], B07[C-I], B07[C-B], B07[C-Si], B07[C-X], B07[N-N], B07[N-O], B07[N-S], B07[N-P], B07[N-F], B07[N-Cl], B07[N-Br], B07[N-I], B07[N-B], B07[N-Si], B07[N-X], B07[O-O], B07[O-S], B07[O-P], B07[O-F], B07[O-Cl], B07[O-Br], B07[O-I], B07[O-B], B07[O-Si], B07[O-X], B07[S-S], B07[S-P], B07[S-F], B07[S-Cl], B07[S-Br], B07[S-I], B07[S-B], B07[S-Si], B07[SX], B07[P-P], B07[P-F], B07[P-Cl], B07[P-Br], B07[P-I], B07[P-B], B07[P-Si], B07[P-X], B07[F-F], B07[F-Cl], B07[F-Br], B07[F-I], B07[F-B], B07[F-Si], B07[F-X], B07[Cl-Cl], B07[Cl-Br], B07[Cl-I], B07[Cl-B], B07[Cl-Si], B07[Cl-X], B07[Br-Br], B07[Br-I], B07[Br-B], B07[Br-Si], B07[Br-X], B07[I-I], B07[I-B], B07[I-Si], B07[I-X], B07[B-B], B07[B-Si], B07[B-X], B07[Si-Si], B07[Si-X], B07[X-X], B08[CC], B08[C-N], B08[C-O], B08[C-S], B08[C-P], B08[C-F], B08[C-Cl], B08[C-Br], B08[C-I], B08[C-B], B08[C-Si], B08[C-X], B08[N-N], B08[N-O], B08[N-S], B08[N-P], B08[N-F], B08[N-Cl], B08[N-Br], B08[N-I], B08[N-B], B08[N-Si], B08[N-X], B08[O-O], B08[O-S], B08[O-P], B08[O-F], B08[O-Cl], B08[O-Br], B08[O-I], B08[O-B], B08[O-Si], B08[O-X], B08[S-S], B08[S-P], B08[S-F], B08[S-Cl], B08[S-Br], B08[S-I], B08[S-B], B08[S-Si], B08[S-X], B08[P-P], B08[P-F], B08[P-Cl], B08[P-Br], B08[P-I], B08[P-B], B08[P-Si], B08[P-X], B08[F-F], B08[F-Cl], B08[F-Br], B08[F-I], B08[F-B], B08[F-Si], B08[F-X], B08[Cl-Cl], B08[Cl-Br], B08[Cl-I], B08[Cl-B], B08[Cl-Si], B08[Cl-X], B08[Br-Br], B08[Br-I], B08[Br-B], B08[Br-Si], B08[Br-X], B08[I-I], B08[I-B], B08[I-Si], B08[I-X], B08[B-B], B08[B-Si], B08[B-X], B08[Si-Si], B08[Si-X], B08[X-X], B09[C-C], B09[C-N], B09[C-O], B09[C-S], B09[C-P], B09[C-F], B09[C-Cl], B09[C-Br], B09[C-I], B09[C-B], B09[C-Si], B09[C-X], B09[N-N], B09[N-O], B09[N-S], B09[N-P], B09[N-F], B09[N-Cl], B09[N-Br], B09[N-I], B09[N-B], B09[N-Si], B09[N-X], B09[O-O], B09[O-S], B09[O-P], B09[O-F], B09[O-Cl], B09[O-Br], B09[O-I], B09[O-B], B09[O-Si], B09[O-X], B09[S-S], B09[S-P], B09[S-F], B09[S-Cl], B09[S-Br], B09[S-I], B09[S-B], B09[S-Si], B09[S-X], B09[P-P], B09[P-F], B09[P-Cl], B09[P-Br], B09[P-I], B09[P-B], B09[P-Si], B09[P-X], B09[F-F], B09[F-Cl], B09[F-Br], B09[F-I], B09[F-B], B09[F-Si], B09[F-X], B09[Cl-Cl], B09[Cl-Br], B09[Cl-I], B09[Cl-B], B09[Cl-Si], B09[Cl-X], B09[Br-Br], B09[Br-I], B09[Br-B], B09[Br-Si], B09[Br-X], B09[I-I], B09[IB], B09[I-Si], B09[I-X], B09[B-B], B09[B-Si],</p>
--	--	--

		<p>B09[B-X], B09[Si-Si], B09[Si-X], B09[X-X], B10[C-C], B10[C-N], B10[C-O], B10[C-S], B10[C-P], B10[C-F], B10[C-Cl], B10[C-Br], B10[C-I], B10[C-B], B10[C-Si], B10[C-X], B10[N-N], B10[N-O], B10[N-S], B10[N-P], B10[N-F], B10[N-Cl], B10[N-Br], B10[N-I], B10[N-B], B10[N-Si], B10[N-X], B10[O-O], B10[O-S], B10[O-P], B10[O-F], B10[O-Cl], B10[O-Br], B10[O-I], B10[O-B], B10[O-Si], B10[O-X], B10[S-S], B10[S-P], B10[S-F], B10[S-Cl], B10[S-Br], B10[S-I], B10[S-B], B10[S-Si], B10[S-X], B10[P-P], B10[P-F], B10[P-Cl], B10[P-Br], B10[P-I], B10[P-B], B10[P-Si], B10[P-X], B10[F-F], B10[F-Cl], B10[F-Br], B10[F-I], B10[F-B], B10[F-Si], B10[F-X], B10[Cl-Cl], B10[Cl-Br], B10[Cl-I], B10[Cl-B], B10[Cl-Si], B10[Cl-X], B10[Br-Br], B10[Br-I], B10[Br-B], B10[Br-Si], B10[Br-X], B10[I-I], B10[I-B], B10[I-Si], B10[I-X], B10[B-B], B10[B-Si], B10[B-X], B10[Si-Si], B10[Si-X], B10[X-X], F01[C-C], F01[C-N], F01[C-O], F01[C-S], F01[C-P], F01[C-F], F01[C-Cl], F01[C-Br], F01[C-I], F01[C-B], F01[C-Si], F01[C-X], F01[N-N], F01[N-O], F01[N-S], F01[N-P], F01[N-F], F01[N-Cl], F01[N-Br], F01[N-I], F01[N-B], F01[N-Si], F01[N-X], F01[O-O], F01[OS], F01[O-P], F01[O-F], F01[O-Cl], F01[O-Br], F01[O-I], F01[O-B], F01[O-Si], F01[O-X], F01[S-S], F01[S-P], F01[S-F], F01[S-Cl], F01[S-Br], F01[S-I], F01[S-B], F01[S-Si], F01[S-X], F01[P-P], F01[P-F], F01[P-Cl], F01[P-Br], F01[P-I], F01[P-B], F01[P-Si], F01[P-X], F01[F-F], F01[F-Cl], F01[F-Br], F01[F-I], F01[F-B], F01[F-Si], F01[F-X], F01[Cl-Cl], F01[Cl-Br], F01[Cl-I], F01[Cl-B], F01[Cl-Si], F01[Cl-X], F01[Br-Br], F01[Br-I], F01[Br-B], F01[Br-Si], F01[Br-X], F01[I-I], F01[I-B], F01[I-Si], F01[I-X], F01[B-B], F01[B-Si], F01[B-X], F01[Si-Si], F01[Si-X], F01[X-X], F02[C-C], F02[C-N], F02[C-O], F02[C-S], F02[C-P], F02[C-F], F02[C-Cl], F02[C-Br], F02[C-I], F02[C-B], F02[C-Si], F02[C-X], F02[N-N], F02[N-O], F02[N-S], F02[N-P], F02[N-F], F02[N-Cl], F02[N-Br], F02[N-I], F02[N-B], F02[N-Si], F02[NX], F02[O-O], F02[O-S], F02[O-P], F02[O-F], F02[O-Cl], F02[O-Br], F02[O-I], F02[O-B], F02[O-Si], F02[O-X], F02[S-S], F02[S-P], F02[SF], F02[S-Cl], F02[S-Br], F02[S-I], F02[S-B], F02[S-Si], F02[S-X], F02[P-P], F02[P-F], F02[P-Cl], F02[P-Br], F02[P-I], F02[P-B], F02[P-Si], F02[P-X], F02[F-F], F02[F-Cl], F02[F-Br], F02[F-I], F02[F-B], F02[F-Si], F02[F-X], F02[Cl-Cl], F02[Cl-Br], F02[Cl-I], F02[Cl-B], F02[Cl-Si], F02[Cl-X], F02[Br-Br], F02[Br-I], F02[Br-B], F02[Br-Si], F02[Br-X], F02[I-I], F02[I-B], F02[I-Si], F02[I-X], F02[B-B], F02[B-Si], F02[B-X], F02[Si-Si], F02[Si-X], F02[X-X], F03[C-C], F03[C-N], F03[C-O], F03[C-S], F03[C-P], F03[C-F], F03[C-Cl], F03[C-Br], F03[C-I], F03[C-B], F03[C-Si], F03[C-X], F03[N-N], F03[N-O], F03[N-S], F03[N-P], F03[N-F], F03[N-Cl], F03[N-Br], F03[N-I], F03[N-B], F03[N-Si], F03[N-X], F03[O-O], F03[O-S], F03[O-P], F03[O-F], F03[O-Cl], F03[O-Br], F03[O-I], F03[O-B], F03[O-Si], F03[O-X], F03[SS], F03[S-P], F03[S-F], F03[S-Cl], F03[S-Br], F03[S-I], F03[S-B], F03[S-Si], F03[S-X], F03[P-P], F03[P-F], F03[P-Cl], F03[P-Br], F03[P-I], F03[P-B], F03[P-Si], F03[P-X], F03[F-F], F03[F-Cl], F03[F-Br], F03[F-I], F03[F-B], F03[F-Si], F03[F-X], F03[Cl-Cl], F03[Cl-Br], F03[Cl-I], F03[Cl-B], F03[Cl-Si], F03[Cl-X], F03[Br-Br], F03[Br-I], F03[Br-B], F03[Br-Si], F03[Br-X], F03[I-I], F03[I-B], F03[I-Si], F03[I-X], F03[B-B],</p>
--	--	---

		<p>F03[B-Si], F03[B-X], F03[Si-Si], F03[Si-X], F03[X-X], F04[C-C], F04[C-N], F04[C-O], F04[C-S], F04[C-P], F04[C-F], F04[C-Cl], F04[C-Br], F04[C-I], F04[C-B], F04[C-Si], F04[C-X], F04[N-N], F04[N-O], F04[N-S], F04[N-P], F04[N-F], F04[N-Cl], F04[N-Br], F04[N-I], F04[N-B], F04[N-Si], F04[N-X], F04[O-O], F04[O-S], F04[O-P], F04[O-F], F04[O-Cl], F04[O-Br], F04[O-I], F04[O-B], F04[O-Si], F04[O-X], F04[S-S], F04[S-P], F04[S-F], F04[S-Cl], F04[S-Br], F04[S-I], F04[S-B], F04[S-Si], F04[S-X], F04[P-P], F04[P-F], F04[P-Cl], F04[P-Br], F04[P-I], F04[P-B], F04[P-Si], F04[P-X], F04[F-F], F04[F-Cl], F04[F-Br], F04[F-I], F04[F-B], F04[F-Si], F04[F-X], F04[Cl-Cl], F04[Cl-Br], F04[Cl-I], F04[Cl-B], F04[Cl-Si], F04[Cl-X], F04[Br-Br], F04[Br-I], F04[Br-B], F04[Br-Si], F04[Br-X], F04[I-I], F04[I-B], F04[I-Si], F04[I-X], F04[B-B], F04[B-Si], F04[B-X], F04[Si-Si], F04[Si-X], F04[X-X], F05[C-C], F05[C-N], F05[C-O], F05[C-S], F05[C-P], F05[C-F], F05[C-Cl], F05[C-Br], F05[C-I], F05[C-B], F05[C-Si], F05[C-X], F05[N-N], F05[N-O], F05[N-S], F05[N-P], F05[N-F], F05[N], Cl, F05[N-Br], F05[N-I], F05[N-B], F05[N-Si], F05[N-X], F05[O-O], F05[O-S], F05[O-P], F05[O-F], F05[O-Cl], F05[O-Br], F05[O-I], F05[O-B], F05[O-Si], F05[O-X], F05[S-S], F05[S-P], F05[S-F], F05[S-Cl], F05[S-Br], F05[S-I], F05[S-B], F05[S-Si], F05[S-X], F05[P-P], F05[P-F], F05[P-Cl], F05[P-Br], F05[P-I], F05[P-B], F05[P-Si], F05[P-X], F05[F-F], F05[F-Cl], F05[F-Br], F05[F-I], F05[F-B], F05[F-Si], F05[F-X], F05[Cl-Cl], F05[Cl-Br], F05[Cl-I], F05[Cl-B], F05[Cl-Si], F05[Cl-X], F05[Br-Br], F05[Br-I], F05[Br-B], F05[Br-Si], F05[Br-X], F05[I-I], F05[I-B], F05[I-Si], F05[I-X], F05[B-B], F05[B-Si], F05[B-X], F05[Si-Si], F05[Si-X], F05[X-X], F06[C-C], F06[C-N], F06[C-O], F06[C-S], F06[C-P], F06[C-F], F06[C-Cl], F06[C-Br], F06[C-I], F06[C-B], F06[C-Si], F06[C-X], F06[N-N], F06[N-O], F06[N-S], F06[N-P], F06[N-F], F06[N-Cl], F06[N-Br], F06[N-I], F06[N-B], F06[N-Si], F06[N-X], F06[O-O], F06[O-S], F06[O-P], F06[O-F], F06[O-Cl], F06[O-Br], F06[O-I], F06[O-B], F06[O-Si], F06[O-X], F06[S-S], F06[S-P], F06[S-F], F06[S-Cl], F06[S-Br], F06[S-I], F06[S-B], F06[S-Si], F06[S-X], F06[P-P], F06[P-F], F06[P-Cl], F06[P-Br], F06[P-I], F06[P-B], F06[P-Si], F06[P-X], F06[F-F], F06[F-Cl], F06[F-Br], F06[F-I], F06[FB], F06[F-Si], F06[F-X], F06[Cl-Cl], F06[Cl-Br], F06[Cl-I], F06[Cl-B], F06[Cl-Si], F06[Cl-X], F06[Br-Br], F06[Br-I], F06[Br-B], F06[Br-Si], F06[Br-X], F06[I-I], F06[I-B], F06[I-Si], F06[I-X], F06[B-B], F06[B-Si], F06[B-X], F06[Si-Si], F06[Si-X], F06[X-X], F07[C-C], F07[CN], F07[C-O], F07[C-S], F07[C-P], F07[C-F], F07[C-Cl], F07[C-Br], F07[C-I], F07[C-B], F07[C-Si], F07[C-X], F07[N-N], F07[N-O], F07[N-S], F07[N-P], F07[N-F], F07[N-Cl], F07[N-Br], F07[N-I], F07[N-B], F07[N-Si], F07[N-X], F07[O-O], F07[O-S], F07[O-P], F07[OF], F07[O-Cl], F07[O-Br], F07[O-I], F07[O-B], F07[O-Si], F07[O-X], F07[S-S], F07[S-P], F07[S-F], F07[S-Cl], F07[S-Br], F07[S-I], F07[SB], F07[S-Si], F07[S-X], F07[P-P], F07[P-F], F07[P-Cl], F07[P-Br], F07[P-I], F07[P-B], F07[P-Si], F07[P-X], F07[F-F], F07[F-Cl], F07[F-Br], F07[F-I], F07[F-B], F07[F-Si], F07[F-X], F07[Cl-Cl], F07[Cl-Br], F07[Cl-I], F07[Cl-B], F07[Cl-Si], F07[Cl-X], F07[Br-Br], F07[Br-I], F07[Br-B], F07[Br-Si], F07[Br-X], F07[I-I], F07[I-B], F07[I-Si], F07[I-X], F07[B-B], F07[B-Si], F07[B-X], F07[Si-</p>
--	--	--

		<p>Si], F07[Si-X], F07[XX],F08[C-C], F08[C-N], F08[C-O], F08[C-S], F08[C-P], F08[C-F], F08[C-Cl], F08[C-Br], F08[C-I], F08[C-B], F08[C-Si], F08[C-X],F08[N-N], F08[N-O], F08[N-S], F08[N-P], F08[N-F], F08[N-Cl], F08[N-Br], F08[N-I], F08[N-B], F08[N-Si], F08[N-X], F08[O-O], F08[OS],F08[O-P], F08[O-F], F08[O-Cl], F08[O-Br], F08[O-I], F08[O-B], F08[O-Si], F08[O-X], F08[S-S], F08[S-P], F08[S-F], F08[S-Cl], F08[S-Br],F08[S-I], F08[S-B], F08[S-Si], F08[S-X], F08[P-P], F08[P-F], F08[P-Cl], F08[P-Br], F08[P-I], F08[P-B], F08[P-Si], F08[P-X], F08[F-F],F08[F-Cl], F08[F-Br], F08[F-I], F08[F-B], F08[F-Si], F08[F-X], F08[Cl-Cl], F08[Cl-Br], F08[Cl-I], F08[Cl-B], F08[Cl-Si], F08[Cl-X],F08[Br-Br], F08[Br-I], F08[Br-B], F08[Br-Si], F08[Br-X], F08[I-I], F08[I-B], F08[I-Si], F08[I-X], F08[B-B], F08[B-Si], F08[B-X], F08[Si-Si], F08[Si-X], F08[X-X], F09[C-C], F09[C-N], F09[C-O], F09[C-S], F09[C-P], F09[C-F], F09[C-Cl], F09[C-Br], F09[C-I], F09[C-B],F09[C-Si], F09[C-X], F09[N-N], F09[N-O], F09[N-S], F09[N-P], F09[N-F], F09[N-Cl], F09[N-Br], F09[N-I], F09[N-B], F09[N-Si], F09[NX],F09[O-O], F09[O-S], F09[O-P], F09[O-F], F09[O-Cl], F09[O-Br], F09[O-I], F09[O-B], F09[O-Si], F09[O-X], F09[S-S], F09[S-P], F09[SF],F09[S-Cl], F09[S-Br], F09[S-I], F09[S-B], F09[S- Si], F09[S-X], F09[P-P], F09[P-F], F09[P-Cl], F09[P-Br], F09[P-I], F09[P-B], F09[-Si],F09[P-X], F09[F-F], F09[F-Cl], F09[F-Br], F09[F-I], F09[F-B], F09[F-Si], F09[F-X], F09[Cl-Cl], F09[Cl-Br], F09[Cl-I], F09[Cl-B],F09[Cl-Si], F09[Cl-X], F09[Br-Br], F09[Br-I], F09[Br-B], F09[Br-Si], F09[Br-X], F09[I-I], F09[I-B], F09[I-Si], F09[I-X], F09[B-B], F09[B-Si],F09[B-X], F09[Si-Si], F09[Si-X], F09[X-X], F10[C-C], F10[C-N], F10[C-O], F10[C-S], F10[C-P], F10[C-F], F10[C-Cl], F10[C-Br],F10[C-I], F10[C-B], F10[C-Si], F10[C-X], F10[N-N], F10[N-O], F10[N-S], F10[N-P], F10[N-F], F10[N-Cl], F10[N-Br], F10[N-I], F10[N-B], F10[N-Si], F10[N-X], F10[O-O], F10[O-S], F10[O-P], F10[O-F], F10[O-Cl], F10[O-Br], F10[O-I], F10[O-B], F10[O-Si], F10[O-X], F10[SS], F10[S-P], F10[S-F], F10[S-Cl], F10[S-Br], F10[S-I], F10[S-B], F10[S-Si], F10[S-X], F10[P-P], F10[P-F], F10[P-Cl], F10[P-Br], F10[P-I],F10[P-B], F10[P-Si], F10[P-X], F10[F-F], F10[F-Cl], F10[F-Br], F10[F-I], F10[F-B], F10[F-Si], F10[F-X], F10[Cl-Cl], F10[Cl-Br], F10[Cl-I], F10[Cl-B], F10[Cl-Si], F10[Cl-X], F10[Br-Br], F10[Br-I], F10[Br-B],F10[Br-Si],F10[Br-X], F10[I-I], F10[I-B], F10[I-Si],F10[I-X],F10[B-B], F10[B-Si], F10[B-X], F10[Si-Si], F10[Si-X], F10[X-X]</p>
5	Ring descriptors	<p>nCIC ,nCIR ,TRS ,Rperim ,Rbrid , MCD, RFD ,RCI , NRS, NNRS, nR03, nR04 , nR05, nR06, nR07 ,nR08 ,nR09 nR10 , nR11, nR12 ,nBnz , ARR , D/Dtr03 ,D/Dtr04 , D/Dtr05,D/Dtr06,D/Dtr07, D/Dtr08 , D/Dtr09 ,D/Dtr10 ,D/Dtr11</p>
6	Connectivity index	<p>X0, X1, X2, X3, X4, X5, X0A, X1A, X2A, X3A, X4A, X5A, X0v, X1v, X2v, X4v, X5v, X0Av, X1Av, X2Av, X3Av ,X4Av, X5Av ,X0sol X1sol, X2sol, X3sol, X4sol, X5sol, XMOD, RDCHI, RDSQ, X1Kup, X1Mad, X1Pe, X1MulPer</p>

7	Molecular properties	Ui, Hy, AMR, TPSA(NO), TPSA(Tot), MLOGP, MLOGP2, ALOGP, ALOGP2, SA _{tot} , SA _{acc} , SA _{don} , Vx, VvdwM, VvdwZAZ, PDI, BLTF96, BLTD48, BLTA96
8	Atom centered fragments	C-001 ,C-002,C-003, C-004, C-005, C-006,,C-007, C-008 C-009, C-010, C-011, C-012, C-013, C-014, C-015, C-016, C-017, C-018, C-019, C-020, C-021, C-022, C-023, C-024, C-025, C-026, C-027, C-028, C-029, C-030, C-031, C-032, C-033, C-034, C-035, C-036, C-037, C-038, C-039, C-040, C-041, C-042 ,C-043, C-044, H-046, H-047 ,H-048, H-049, H-050, H-051, H-052 ,H-053, H-054, H-055, O-056, O-057, O-058, O-059, O-060, O-061, O-062, O-063, Se-064, Se-065, N-066, N-067, N-068, N-069, N-070, N-071, N-072, N-073, N-074, N-075, N-076, N-077, N-078, N-079, F-081, F-082 , F-083, F-084, F-085, Cl-086, Cl-087 ,Cl-088, Cl-089, Cl-090, Br-091, Br-092, Br-093, Br-094, Br-095 ,I-096, I-097, I-098, I-099, I-100, F-101, Cl-102, Br-103, I-104, S-106, S-107, S-108, S-109, S-110 ,Si-111, B-112, P-115 ,P-116, P-117, P-118 ,P-119 ,P-120

Table 3.4: Extended Topochemical Atom (ETA) indices (obtained from PaDel software)

Sl. No.	Definition	Significance
1	$\Delta\alpha_A = \left\langle \frac{\sum \alpha - [\sum \alpha]_R}{N_V} \right\rangle$	A measure of count of non-hydrogen heteroatoms [N _V stands for total number of atoms excluding hydrogens]
2	$\Delta\alpha_B = \left\langle \frac{[\sum \alpha]_R - \sum \alpha}{N_V} \right\rangle$	A measure of count of hydrogen bond acceptor atoms and/or polar surface area
3	$\varepsilon_1 = \frac{\sum \varepsilon}{N}$	A measure of electronegative atom count [N stands for total number of atoms including hydrogens]
4	$\varepsilon_2 = \frac{\sum \varepsilon_{EH}}{N_V}$	A measure of electronegative atom count [EH stands for <i>excluding hydrogens</i>]
5	$\varepsilon_3 = \frac{[\sum \varepsilon]_R}{N_R}$	[R stands for reference alkane]
6	$\varepsilon_4 = \frac{[\sum \varepsilon]_{SS}}{N_{SS}}$	[SS stands for saturated carbon skeleton]

7	$\varepsilon_5 = \frac{\sum \varepsilon_{EH} + \sum \varepsilon_{XH}}{N_V + N_{XH}}$	[XH stands for those hydrogens which are connected to a heteroatom]
8	$\Delta\varepsilon_A = \varepsilon_1 - \varepsilon_3$	A measure of contribution of unsaturation and electronegative atom count
9	$\Delta\varepsilon_B = \varepsilon_1 - \varepsilon_4$	A measure of contribution of unsaturation
10	$\Delta\varepsilon_C = \varepsilon_3 - \varepsilon_4$	A measure of contribution of electronegativity
11	$\Delta\varepsilon_D = \varepsilon_2 - \varepsilon_5$	A measure of contribution of hydrogen bond donor atoms
12	$\psi = \frac{\alpha}{\varepsilon}$	A measure of hydrogen bonding propensity of the atoms
13	$\psi_1 = \frac{\sum \alpha}{\sum \varepsilon]_{EH}} = \frac{\sum \alpha / N_V}{\varepsilon_2}$	A measure of hydrogen bonding propensity of the molecules and/or polar surface area
14	$\Delta\psi_A = \langle 0.714 - \psi_1 \rangle$	A measure of hydrogen bonding propensity of the molecules
15	$\Delta\psi_B = \langle \psi_1 - 0.714 \rangle$	A measure of hydrogen bonding propensity of the molecules
16	$\Delta\beta = \sum \beta_{ns} - \sum \beta_s$	A measure of relative unsaturation content
17	$\Delta\beta' = \frac{\Delta\beta}{N_V}$	A measure of relative unsaturation content
18	$\sum \beta_{ns(\delta)}$	A measure of lone electrons entering into resonance
19	$\sum \beta'_{ns(\delta)} = \frac{\sum \beta_{ns(\delta)}}{N_V}$	A measure of lone electrons entering into resonance

3.2.1.1. Integral additive descriptors

The modes of action (generally 4 major types) for some mixture of conventional organic compounds were defined as **a**) simple additive, i.e., the combined toxic response is equal to the total of the single chemical toxicity, **b**) greater than synergism or additive that means that the joint effect is more than the sum of the individual chemicals toxicity, **c**) the total toxicity

is less than the overall individual chemical toxicity, i.e., less than additive or partial additive, **d)** independent or no interaction which means the combined toxic effect is equal to that caused by the component with highest toxicity. Here, each individual component contributes additively to the toxicity and its contribution is proportional to the individual component mole fraction in the mixture. Summation of the concentration of the individual components in the mixture was carried out after multiplying each with a scaling factor that indicates the contribution of property of the individual components (C_i) and hence the sum of the concentration (a) is the property of mixture based NPs (C_{mix}) as shown in **Eq.3.1**.

$$C_{mix} = \sum_{i=1}^n aC_i \quad (3.1)$$

The individual component can be expressed as a set of 2D and 3D descriptors in the framework of integral additive scheme and the descriptor is expressed as mole weighted average and mole fraction of each component as follows in **Eq.3.2**:

$$D_{mix} = R_1D_1 + R_nD_n \dots \dots \quad (3.2)$$

Where, D_{mix} corresponds to the mixture descriptor, R_1 and R_n represent the mole fraction of the individual component in the mixture, and D_1 and D_n stands for descriptors of each component in the mixture. The used descriptors are listed in the **Table 3.4**.

Table 3.4: The lists of periodic table descriptors used in the development of models

Sl.no	Descriptors	description
1	mol wt	Molecular weight is the mass of a molecule. It is calculated as the sum of the atomic weights of each constituent element multiplied by the number of atoms of that element in the molecular formula
2	electron affinity (kJ/mol)	The electron affinity (E_{ea}) of an atom or molecule is defined as the amount of energy released or spent when an electron is added to a neutral atom or molecule in the

		gaseous state to form a negative ion.
3	Amount of Ag	Amount of the noble metal silver
4	Amount of Au	Amount of the noble metal gold
5	Amount of Pt	Amount of the noble metal platinum
6	valence of Au	Valence of gold
7	valence of Ag	Valence of silver
8	valence of Pt	Valence of platinum
9	BET surface area	
10	molar surface area	
11	Mol_wt	Molecular weight is the mass of a molecule. It is calculated as the sum of the atomic weights of each constituent element multiplied by the number of atoms of that element in the molecular formula.
12	surface vol ratio	Surface to volume ratio of the noble metal
13	Nmetal	Number of metal
14	χ (electronegativity) Pauling scale	Electronegativity, symbol χ , is a chemical property that describes the tendency of an atom to attract a shared pair of electrons (or electron density) towards itself. An atom's electronegativity is affected by both its atomic number and the distance at which its valence electrons reside from the charged nucleus. The higher the associated electronegativity number, the more an element

		or compound attracts electrons towards it.
15	Z _{metal} (atomic no)	The atomic number or proton number (symbol Z) of a chemical element is the number of protons found in the nucleus of an atom. It is identical to the charge number of the nucleus. The atomic number uniquely identifies a chemical element. In an uncharged atom, the atomic number is also equal to the number of electrons.
16	Valence	The valence or valency of an element is a measure of its combining power with other atoms when it forms chemical compounds or molecules.
17	Atomic radius (pm)	The atomic radius of a chemical element is a measure of the size of its atoms, usually the mean or typical distance from the center of the nucleus to the boundary of the surrounding cloud of electrons.
18	covalent radius (pm)	The covalent radius, r_{cov} , is a measure of the size of an atom that forms part of one covalent bond. It is usually measured either in picometer (pm) or angstroms (Å), with $1 \text{ \AA} = 100 \text{ pm}$.

19	van der waal radius (pm)	The van der Waals radius, r_w , of an atom is the radius of an imaginary hard sphere representing the distance of closest approach for another atom.
20	radius	Radius of the particular atom
21	hardness	
22	1st ionization (kJ/mol)	The first ionization energy is the energy required to remove one mole of the most loosely held electrons from one mole of gaseous atoms to produce 1 mole of gaseous ions each with a charge of 1+.
23	2nd ionization (kJ/mol)	It is the energy needed to remove a second electron from each ion in 1 mole of gaseous 1+ ions to give gaseous 2+ ions
24	thermal conductivity (W/(m·K))	Thermal conductivity (often denoted k , λ , or κ) is the property of a material to conduct heat.
25	M.P (K)	Melting point of the noble metal
26	Z_v _metal	Valence electron of metal
27	density (g/cm ³)	Density is a measurement that compares the amount of matter an object has to its volume
28	electrical resistivity ($n\Omega \cdot m$)	Electrical resistivity (also known as resistivity, specific electrical resistance, or volume resistivity) is a fundamental property of a material that quantifies how strongly that material opposes the flow of electric current

29	valance electron potential (-eV)	Provides a quantitative indication of elements reactivity and is based on the charge of the valence electrons and the ionic radius
30	electron work function (eV)	The smallest amount of photonic energy necessary to remove an electron from the boundary of an element
31	Electrochemical Equivalent (g/amp-hr)	The electrochemical equivalent, sometimes abbreviated Eq, of a chemical element is the mass of that element (in grams) transported by 1 coulomb of electric charge
32	Heat of Fusion (kJ/mol)	Heat of fusion is the change in its enthalpy resulting from providing energy, typically heat, to a specific quantity of the substance to change its state from a solid to a liquid, at constant pressure.
33	$(Z-Z_v)/Z_v_{\text{metal}}$	Core environment of metal defined by the ratio of the number of core electrons to the number of valence electrons

3.2.2. Pretreatment of descriptors: thinning of the pool

In order to obviate the impact of redundant and noisy variables various pretreatment operations are employed. The descriptors are required to capture chemical information of the employed compounds and portray the change in chemical structures e.g., functional groups, branching etc. accordingly in a dataset. Hence, from a total pool of calculated descriptors those with constant variance and or intercorrelation features can be omitted leaving the most contributory ones in the modified pool. In our studies, we have employed variance cut of 0.0001 and intercorrelation cut off of 0.95 as a thinning strategy.

3.2.3. Division of the dataset: selection of the training set and test set

Division of the whole data presents a crucial step for developing an acceptable QSAR models. The dataset can be divided into training and test sets by using division methods such as Euclidean distance (diversity-based), Kennard Stone, k-means clustering and sorted response. The splitting of the dataset was done in such a way that the training set would capture all the information of the dataset enabling correct predictions for the test set compounds from the corresponding QSAR model. In our studies, we have employed k-means clustering method as the dataset division strategy. The central concept is to categorize the compounds into different groups or clusters based on their chemical nature followed by the selection of representative compounds into training and test sets from each cluster. The k-means clustering represent non-hierarchical clustering formalism, and is useful with the known number of samples in each cluster. Here, an unsupervised method is implemented that gives rise to k clusters based on k number of centroids. The numbers of centroids are arranged maintaining the longest distance from each other and the samples associated to them are computed. With no points remaining, the positions of the centroids are recalculated and repeated until the centroids do not move. By this way, an optimal numbers of k clusters are obtained from which predefined fraction of compounds can be selected into training and test sets.

3.2.3.1. k-Means clustering

k-Means is one of the simplest unsupervised learning algorithms [83] that solves the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed *a priori*. The main idea is to define ‘ k ’ centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. This algorithm aims at minimizing an objective function known as squared error function given by (Eq.3.3) [84].

$$(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (||x_i - v_j||)^2 \quad (3.3)$$

Where, $||x_i - v_j||$ is the Euclidean distance between x_i and v_j
 c_i is the number of data points in the i^{th} cluster

'c' is the number of cluster centers

The steps involved in the *k*-means clustering method are as follows:

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ be the set of data points and $V = \{v_1, v_2, \dots, v_c\}$ be the set of centers.

- (i) Randomly select 'c' cluster centers.
- (ii) Calculate the distance between each data point and cluster centers.
- (iii) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
- (iv) Recalculate the new cluster center using (Eq.3.4):

$$v_i = \left(\frac{1}{c_i}\right) \sum_{j=1}^{c_i} x_j \quad (3.4)$$

Where, 'c_i' represents the number of data points in *i*th cluster.

- (v) Recalculate the distance between each data point and new obtained cluster centers.
- (vi) If no data point was reassigned then stop, otherwise repeat from step (iii).

3.2.4. Selection of feature

Selection of variables is a very important step in QSPR analysis. Relatively a large number of predictor variables are computed for developing predictive correlation, the final model is expected to be derived from selected variables with most suitable chemical diagnostic power with respect to the response being addressed. Hence, different chemometric tools play a pivotal role in selecting the suitable descriptors. The number of predictor variables should be at least one fifth to the number of compounds employed for developing the model. In our work, we have employed stepwise selection method, which is discussed below.

3.2.4.1. Stepwise regression method

Stepwise regression is a type of multiple linear regression equation made step by step which is altered by adding or removing a predictor variable. Forward selection and backward elimination are two parts of stepwise regression method [85]. Forward selection starts with

no variable and then ‘statistically significant’ variables are included one by one. In case of backward elimination, initially, all the candidate variables are selected and then deleting statistically insignificant variables one by one. The objective function of the selection in stepwise regression may be “F-for-inclusion”, also known as “stepping criteria”. The F-value is square of t value of incoming variable which signify the regression coefficient. In stepwise regression process, a multiple term linear regression equation is built up using a “stepping criteria” also known as “Fisher criteria”[86]. The F-value used for inclusion or exclusion of a variable in the stepwise regression process is a test for partial regression coefficient and it is obtained by dividing the difference between reductions of sum of squares with and without the variable being included or excluded with error mean square of the equation [87]. The “stepping criteria” or “Fisher criteria” was fixed at F=4 to enter and F=3.9 to remove [88] because at this value of the F-criterion, the descriptors are considered to be significant at the 95% confidence level. A limitation of the stepwise regression search approach is that it presumes that there is a single 'best' subset of X variables and seeks to identify it.

3.2.5. Employed statistical modeling techniques

In our present study, we have used regression based formalisms to develop predictive models on different property endpoints. In regression based formalism we aimed to predict the exact property value of the compounds. Here, the regression based QSPR formalism involves multiple linear regression (MLR) and partial least squares (PLS) techniques.

3.2.5.1. Multiple linear regressions (MLR)

Regression analysis is the method for establishing mathematical relationship between one or more response variables. Linear regression is an approach to develop a statistical relationship between a scalar variable Y (commonly termed as) and one or more variables denoted X (dependent variable). In multiple linear regressions, one dependent variable is correlated with more than one independent ones. The response variable is assumed to be a linear function of the model parameters. The general form of a MLR equation can be represented as in **Eq.3.5**.

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots + a_nx_n \quad (3.5)$$

As we are talking in terms of mathematical, this equation is better explained in terms of variables. Here, y represents respond being modeled i.e. activity, toxicity, property while $X_1, X_2, X_3, \dots, X_n$ represents the independent variables denoting the physiological properties in terms of numerical quantities and $a_0, a_1, a_2, a_3, \dots, a_n$ stands for the contributes of individual descriptors with a_0 as constant term [89].

3.2.5.2. Partial least squares (PLS) techniques

PLS is a generalization of regression which is particularly suited when the matrix of predictors has more variables than observations, and when there is multicollinearity among X variables. PLS is used to find the fundamental relations between X and Y matrices using a latent variable approach to model the covariance structures in these two spaces. Application of PLS allows the construction of larger QSAR equations while still avoiding overfitting and eliminating most variables. PLS is statistically more robust than MLR because standard regression will fail in such cases [90]. However, MLR is more straightforward in calculation than PLS as the former does not require calculation of any latent variables and optimization of the number of components. PLS is normally used in combination with cross-validation to obtain the optimum number of components. This ensures that the QSAR equations are selected based on their ability to predict the data rather than to fit the data [91]. In the present study, leave-one-out (LOO) cross validation method is employed to select optimum number of components [92]. Based on the standardized regression coefficients, the variables with smaller coefficients have been removed from the PLS regression, until there was no further improvement in Q^2 value, irrespective of the components.

3.2.5.3. Intelligent consensus predictor (ICP) [73]: This software was used to judge the performance of consensus predictions and compares them with the prediction quality obtained from the individual (MLR) models based on MAE based criteria (95%). It is obvious that a single model might not be equally useful in prediction for the whole test set compounds which means one QSAR model may be the best model for prediction of a test compound while other model may be the best predictor for another test compounds. For this reason, we have selected five models (M1-M5) in case of dataset containing 59 organic contaminants and performed consensus prediction using “Intelligent consensus predictor”

tool to explore whether the quality of predictions of test set compounds can be enhanced through an “intelligent” selection of multiple MLR models.

3.2.6. Computation of different statistical metrics for assessing model quality

3.2.6.1. Quality measures in fitting of a QSPR model

- **Squared correlation coefficient (R^2):** This parameter is termed as the determination coefficient or squared correlation coefficient. The squared correlation coefficient of a model can be obtained from the following equation **Eq.3.6**:

$$R^2 = 1 - \frac{\sum (Y_{obs} - Y_{cal})^2}{\sum (Y_{obs} - Y_{train})^2} \quad (3.6)$$

The R^2 statistic represents the ratio of the regression variance to the original variance where the former is determined using the original variance minus the variance around the line of regression. The R^2 bears a value between zero (no correlations) to one (perfect correlation). A model possessing a value of R^2 more than 0.8 can be considered to elicit acceptable correlation while the quality enhancing with the increasing value of R^2 until it reaches a maximum value of unity (which is unusual in real cases). Y_{obs} and Y_{calc} are the respective observed and calculated values of the response variable and is their mean value. R^2 gives a measure of explained variance. Each additional X variable added to a model increases R^2 . The prime drawbacks of the R^2 parameter lies in the facts that it does not provide any information on whether: (i) the independent variables are a true cause of the changes in the dependent variable, (ii) the correct regression was used, (iii) the most appropriate set of independent variables has been chosen, (iv) the model might be improved by using transformed versions of the existing set of independent variables and (v) whether any collinearities exists in the data or not.

- **Adjusted R^2 or R_a^2 :**

$$R_a^2 = \frac{(n-1) \times R^2 - P}{n-p-1} \quad (3.7)$$

Adjusted R_a^2 (Eq.3.7) is a modified version of the determination coefficient and is also known as the explained variance. The R_a^2 parameter incorporates the information of the number of samples and the independent variables used in model, and can be defined as follows [93]. Here, R^2 is the determination coefficient of a QSAR model comprising p number of predictor variables and n number of samples. Hence, instead of using only the initial observed (i.e., experimental) and final predicted response values, R_a^2 considers information on the model history in terms of the number of descriptors and number of chemicals used to develop the model (i.e., training set chemicals). The R_a^2 penalizes the R^2 value of a model containing too many independent variables compared to the total number of compounds. The R_a^2 improves only if the addition of a new term enhances the model quality avoiding chance. The R_a^2 value usually is less than the corresponding R^2 value.

- **Standard error of estimate (s):** The error in the estimation of individual activity values of the compounds under study using the MLR method can be quantified based on their residual data. The standard error of estimate (SEE or s) for the residuals is calculated by taking the root-mean square of the residuals. The standard error of the estimate is a measure of the accuracy of fitting. Lower values of SEE correspond to improved model acceptability.

$$s = \sqrt{\frac{\sum (Y_{obs} - Y_{calc})^2}{n - p - 1}} \quad (3.8)$$

In Eq. 3.8, Y_{obs} and Y_{calc} are the actual and estimated scores respectively, while n is the number of scores and p is the number of descriptors.

- **F-value:** F-value (Eq.3.9) is called the variance ratio and is defined

$$F = \frac{\frac{\sum (Y_{calc} - \bar{Y})^2}{p}}{\frac{\sum (Y_{obs} - Y_{calc})^2}{n - p - 1}} \quad (3.9)$$

3.2.6.2. Validation strategies

Both internal and external validation statistics constitute the primary methods for validation of the developed QSPR models. Both the methods have been widely used by different groups of researchers for assessing the predictive ability of the developed model. Another method employs fitting of the dependent X matrix to randomized response parameters. Several metrics are used to check the predictivity of the QSPR models. For the validation of QSPR models, three strategies are primarily adopted: (i) internal validation using the training set molecules and (ii) external validation based on the test set compounds.

3.2.6.2.1. Validation metrics for Training set

3.2.6.2.1.1. Q^2 or Q^2_{int} : The models developed from the training set by using stepwise regression or genetic methods have been subjected to internal validation by means of calculating leave-one-out *cross-validation* R^2 (Q^2) and *predicted residual sum of squares* (*PRESS*) [94] and the acceptable models have been further processed for the prediction of toxicity and/or property of the test set compounds. Cross-validated correlation coefficient R^2 ($\text{LOO-}Q^2$) is calculated according to the formula (Eq.3.10):

$$Q^2 = 1 - \frac{\sum (Y_{\text{obs}(\text{training})} - Y_{\text{pred}(\text{training})})^2}{\sum (Y_{\text{obs}(\text{training})} - \bar{Y}_{\text{training}})^2} \quad (3.10)$$

Here $Y_{\text{obs}(\text{training})}$, $Y_{\text{pred}(\text{training})}$, and $\bar{Y}_{\text{training}}$ are the observed, predicted and the average value of the response variable of the training set. In this technique, one compound is omitted from the data set at random in each cycle and then model is built using the rest of the compounds. The model thus formed in this way is used for the prediction of activity of the omitted compound. The process is iterated until all the compounds are eliminated once. On the basis of the predicting ability of the model, the cross-validated R^2 (Q^2) for the model is determined. Acceptable value of Q^2 is 0.5 with a maximum value of 1.0 and hence more the value is closer to 1, more will be the internal predictivity of the model.

3.2.6.2.1.2. $r_m^2(\text{LOO})$: It was shown that [95] squared cross-validated correlation coefficient alone might not indicate the true predictive capability of a model and hence a modified r^2

($r_m^2(\text{LOO})$) term was used to indicate the leave-one-out prediction capacity of the model for the training set compounds. The parameter $r_m^2(\text{LOO})$ is defined as in (Eq.3.11):

$$r_m^2(\text{LOO}) = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (3.11)$$

where r^2 and r_0^2 are the squared correlation coefficients between the observed and LOO predicted values of the training set compounds with and without intercept respectively. The value of $r_m^2(\text{LOO})$ should be greater than 0.5 for an acceptable model.

3.2.6.2.1.3. Root mean square error in prediction for training set ($rmsep_{\text{int}}$): This parameter suggests that it is possible to determine the internal predictive ability of the training set compounds simply by taking the square root of the squared difference between the observed and predicted response value divided by the number of compounds in the training set [96]. Mathematically (Eq.3.12):

$$rmsep_{\text{int}} = \sqrt{\frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{n_{\text{int}}}} \quad (3.12)$$

where n_{int} is the number of compounds present in the training set and Y_{obs} and Y_{pred} corresponds to the corresponding observed and LOO predicted response value. It should have a minimum value.

3.2.6.2.1.4. Golbraikh and Tropsha criteria

Golbraikh and Tropsha [97] defined a set of criteria to be followed in order to ascertain the external predictive potential of a QSAR model. As we can see that the basic objective of model validation is to determine how close the observed i.e., experimental values are to the corresponding predicted ones. Hence, the simple correlation coefficient between the observed (y) and predicted (\hat{y}) response apparently should give a value of 1 in an ideal case. In this situation, if a regression line is drawn all the points will be located on the line which passes through the origin point (0, 0) in a Cartesian plane. However, in real cases, deviation occurs and the best fitted line poses a definite intercept value. It may be here noted that the plots of

experimental versus fitted or fitted versus experimental response are not equivalent [98] Golbraikh and Tropsha [97] proposed that regression of observed (y) against predicted (\hat{y}) or predicted (\hat{y}) against observed (y) response through the origin must be determined and the corresponding slopes k or k' of the regression lines should be close to unity. This process is known as regression through origin (RTO) method, where a regression line is forcefully passed through the origin point (0, 0) and the corresponding regression lines can be presented as $y^{r_0} = k\hat{y}$ and $\hat{y}^{r_0} = k'y$. The slopes k and k' can be defined as follows in **Eq. 3.13** and **Eq.3.14**:

$$k = \frac{\sum y_i \hat{y}_i}{\sum \hat{y}_i^2} \quad (3.13)$$

and

$$k' = \frac{\sum y_i \hat{y}_i}{\sum y_i^2} \quad (3.14)$$

Golbraikh and Tropsha [99] calculated the determination coefficient values r_0^2 and $r_0'^2$ of the regression lines passing through origin (y against \hat{y} or \hat{y} against y) and, argued that these values should be close to the value of the normal R^2 of the model in case of good predictivity. The r_0^2 and $r_0'^2$ represent the squared correlation coefficient between the observed and predicted response values with and without intercept respectively and can be defined as follows in **Eq. 3.15** and **Eq.3.16**:

$$r_0^2 = 1 - \frac{\sum (\hat{y}_i - y_i^{r_0})^2}{\sum (\hat{y}_i - \bar{\hat{y}})^2} \quad (3.15)$$

$$r_0'^2 = 1 - \frac{\sum (y_i - \hat{y}_i^{r_0'})^2}{\sum (\hat{y}_i - \bar{y})^2} \quad (3.16)$$

Here, \bar{y} and $\bar{\hat{y}}$ refers to the respective mean values of the observed and predicted response data. A set of conditions for model acceptability was proposed by Golbraikh and Tropsha and are summarized below.

- a) $Q_{(LOO)}^2 > 0.5$
 b) $R_{test}^2 > 0.6$
 c) $\frac{r^2 - r_0^2}{r^2} < 0.1$ and $0.85 \leq k \leq 1.15$ or $\frac{r^2 - r_0'^2}{r^2} < 0.1$ and $0.85 \leq k' \leq 1.15$
 d) $|r_0^2 - r_0'^2| < 0.3$

Here, $Q_{(LOO)}^2$ is for the training set only while rest of the parameters correspond to test set chemicals.

3.2.6.2.1.5. The r_m^2 metrics

Using the concept of regression through origin approach, Roy et. al (2009) introduced a new parameter r_m^2 or modified r^2 that penalizes the R^2 value of a model with respect to an ideal condition [100]

The r_m^2 metric can be defined as follows in Eq 3.17 and Eq.3.18:

$$r_m^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0^2)}\right) \quad (3.17) \quad r_m'^2 = r^2 \times \left(1 - \sqrt{(r^2 - r_0'^2)}\right) \quad (3.18)$$

where, r^2 is the squared correlation coefficient value between observed and predicted response values, and r_0^2 and $r_0'^2$ are the respective squared correlation coefficients when the regression line is passed through the origin by interchanging the axes. Roy and co-workers [101-102] further defined average and difference of the two r_m^2 metric values (i.e., r_m^2 and $r_m'^2$) to be used as the acceptable criteria to judge the predictive ability of a model as follows in Eq.3.19 and Eq.3.20.

$$\overline{r_m^2} = \frac{(r_m^2 + r_m'^2)}{2} > 0.5 \quad (3.19) \quad \Delta r_m^2 = |r_m^2 - r_m'^2| < 0.2 \quad (3.20)$$

The r_m^2 metrics can not only be computed for the test set compounds ($r_m^2_{(test)}$) to judge external predictivity, but it can also be used to determine the internal predictivity of the model using the training set. In the latter case, leave-one-out predicted values ($r_m^2_{(LOO)}$) of the training set observations are used against their observed response. Furthermore, Roy et al. [101] also reported the use of the r_m^2 metric in characterizing the overall predictive capability of the model by using leave-one-out predicted values for the training set and equation (i.e., model) based predicted values for the test set together against their corresponding observed

response (r_m^2 (overall)). Later, a rank based r_m^2 [102] as well as a scaled [103] version of the r_m^2 metric was introduced by the same authors' group and these have been used in this present study.

3.2.6.2.1.6. MAE based criteria

In a recent study, Roy et al. [104] have shown that the conventional Q^2 based external validation metrics ($Q^2_{ext(F1)}$, $Q^2_{ext(F2)}$, $Q^2_{ext(F3)}$) often provide biased judgment of model predictivity since such metrics are influenced by factors such as response range and distribution of data. Here, the authors have defined a set of criteria using simple 'mean absolute error' (MAE) and the corresponding standard deviation (σ) measure of the predicted residuals to judge the external predictivity of the models. Note that,

$MAE = \frac{1}{n} \times \sum |Y_{obs} - Y_{pred}|$, where Y_{obs} and Y_{pred} are the respective observed and predicted response values of the test set comprising n number of compounds. The response range of training set compounds has been employed here to define the threshold values. Furthermore, the authors have proposed application of the 'MAE based criteria' on 95% of the test set data by removing 5% data with high predicted residual values precluding the possibility of any outlier prediction. The criteria are described below.

i) Good predictions: The criteria for good predictions is as follows:

$$MAE \leq 0.1 \times \text{training set range AND } (MAE + 3\sigma) \leq 0.2 \times \text{training set range}$$

In simpler terms, an error of 10% of the training set range should be acceptable while an error value more than 20% of the training set range may be considered as high.

ii) Bad predictions: The predictions considered as bad can be defined using the following criteria:

$$MAE > 0.15 \times \text{training set range OR } (MAE + 3\sigma) > 0.25 \times \text{training set range}$$

Here, a value of MAE more than 15% of the training set range is considered high while an error more than 25% of the training set range is judged as very high.

The predictions which do not fall under either of the above two conditions may be considered as of moderate quality. The above criteria should be applied for judging the quality of test set predictions when the number of data points is at least 10 (statistical reliability) and there is no systematic error in model predictions (statistical applicability).

3.2.6.2.2. Validation metric for Test set

3.2.6.2.2.1. R^2_{pred} or $Q^2_{ext(F1)}$

After the prediction of toxicity and/or property of the test set compounds, this parameter was calculated. It can be defined as in **Eq. 3.21**:

$$R^2_{pred} = Q^2_{ext(F1)} = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{\sum (Y_{obs(test)} - \bar{Y}_{training})^2} \quad (3.21)$$

where $Y_{obs(test)}$ is the observed activity of the test set compounds, $Y_{pred(test)}$ is the predicted activity of the test

set compounds and $\bar{Y}_{training}$ corresponds to the mean of observed activity of the training set compounds. R^2_{pred} value for an acceptable model should be greater than 0.5 (maximum value 1).

3.2.6.2.2.2. $Q^2_{ext(F2)}$

This function as a metric for external set validation was described in the paper of Hawkins [105] and can be calculated as in **Eq. 3.22**:

$$Q^2_{ext(F2)} = 1 - \frac{\sum (Y_{obs(test)} - Y_{pred(test)})^2}{\sum (Y_{obs(test)} - \bar{Y}_{test})^2} \quad (3.22)$$

The only notable difference from $Q^2_{ext(F1)}$ is that, in Equation 3.77 the average value of external or test set is used in the denominator instead the internal or training set average value.

Both these functions $Q_{\text{ext (F1)}}^2$ and $Q_{\text{ext (F2)}}^2$ were compared and discussed by Schuurmann *et al.* [104].

3.2.6.2.2.3. $Q_{\text{ext (F3)}}^2$

This function was described by Consonni *et al.* [107] and is defined as (3.23):

$$Q_{\text{ext (F3)}}^2 = 1 - \frac{\sum (Y_{\text{obs (test)}} - Y_{\text{pred (test)}})^2 / n_{\text{test}}}{\sum (Y_{\text{obs (training)}} - \bar{Y}_{\text{training}})^2 / n_{\text{training}}} \quad (3.23)$$

Since the terms for summation in the numerator deals totally with the test set values and the denominator with training set values, division with n_{test} and n_{training} of the numerator and denominator summation expression respectively makes the two squares comparable. The threshold value of acceptance for all the three parameters $Q_{\text{ext (F1)}}^2$, $Q_{\text{ext (F2)}}^2$ and $Q_{\text{ext (F3)}}^2$ is 0.5.

3.2.6.2.2.4. $r_{m^2}^2(\text{test})$

For test set compounds, $r_{m^2}^2(\text{test})$ has been determined which penalizes a model for large differences between observed and predicted values of the test set compounds. The formula is (Eq.3.24):

$$r_{m^2}^2(\text{test}) = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (3.24)$$

r^2 and r_0^2 are the squared correlation coefficients between the observed and predicted values of the test set compounds with and without intercept respectively.

3.2.6.2.2.5. Root mean square error in prediction for test set ($\text{rmsep}_{\text{ext}}$)

We have also calculated the $\text{rmsep}_{\text{ext}}$ parameter as in Eq. 3.25 for the evaluation of external predictive ability of a model as follows [108]

$$\text{rmsep}_{\text{ext}} = \sqrt{\frac{\sum (Y_{\text{obs}} - Y_{\text{pred}})^2}{n_{\text{ext}}}} \quad (3.25)$$

Where n_{ext} represents the number of training set compounds, and Y_{obs} and Y_{pred} corresponds to the observed and predicted activity of the test set compounds respectively. It should have a minimum value. The rmsep value for test and training set depends on the scale of the response activity and therefore comparison makes no sense when a model is compared to another modeling a different activity [99].

3.2.6.2.3. Validation metric for overall set

For the purpose of determination of an overall validation strength of a model, we have calculated the overall r_m^2 metric between the observed toxicity and/or property value of a dataset and the calculated and predicted value of the training and test set respectively. The formula is represented in **Eq.3.26**:

$$r_{m(overall)}^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right) \quad (3.26)$$

The r_m^2 metric has been developed by the present authors' group and has been extensively used by them [100-102].

3.2.6.3. Y-randomization:

The relationships between the response variable and the descriptors can be checked for further statistical significance by randomization test (Y-randomization) of the models. The method can be executed in two ways namely:

- i) Process randomization and
- ii) Model randomization

In process randomization, random scrambling of the dependent response variables is performed accompanied with fresh selection of variables from the whole descriptor matrix and in model randomization scrambling or randomization of the response variable is performed within the descriptors present in an existing model. We have performed process as well as the model randomization of the genetic models. A parameter was proposed by Roy and Paul [109] named R_p^2 that penalises the model R^2 for a small difference between squared mean correlation coefficient (R_r^2) of randomized models and squared correlation coefficient (R^2) of the non-randomized model and was defined as in **Eq.3.27**:

$$R_p^2 = R^2 \times \sqrt{R^2 - R_r^2} \quad (3.27)$$

and the acceptable value of R_p^2 was proposed to be greater than or at least equal to 0.5. Later a correction for this parameter has been suggested by Todeschini [110] and the rebuilt formula is as follows in **Eq 3.28**:

$${}^cR_p^2 = R \times \sqrt{R^2 - R_r^2} \quad (3.28)$$

We have used the new parameter ${}^cR_p^2$ which should be above 0.5 for a good model.

3.2.6.4. Determination of model applicability domain (AD)

Applicability domain (AD) of a QSAR model can be described as the theoretical region in the chemical space defined by the chemical as well as response attributes of the model [111-114]. A definite domain of applicability enables reliability of predictive performance of a model. In other words, any QSPR model possesses a defined theoretical domain within which it can provide reliable predictions of other chemicals not used in developing the model. It is not feasible to develop a single model that can contain the chemical information of the whole universe, and accordingly QSPR models are characterized by different domains. . The applicability domain [115] is a theoretical region in chemical space, defined by the model descriptors and modeled response. When a compound is highly dissimilar to all compounds of the modeling set, reliable prediction of its property is unlikely. The concept of AD was used to avoid such an unjustified extrapolation of property predictions. Here we have applied both Leverage approach and Distance to model in X-space (DModX) approach for verifying the applicability domain of the best model developed from this study.

3.2.6.5. Model validation based on OECD guidelines

To authenticate the applicability of the developed QSPR models and to judge the reliability of the predictions made, the models were further analyzed based on the OECD guidelines [116]. Thus, the QSPR models developed in this work were validated based on these five

guidelines laid down by the OECD. The compliance of the developed models with the OECD guidelines for applicability in regulatory purposes was assessed as follows:

Principle 1 (a defined endpoint):

The response parameter modeled in the present work for the two different datasets were measured under similar conditions. Thus the QSAR models were developed in accordance with the 1st OECD principle.

Principle 2 (an unambiguous algorithm):

Various chemometric tools based on specific algorithms were employed for the calculation of the different categories of descriptors and subsequent QSPR model development using specific software packages. Thus the model development pathway employed for the present studies follow a definite algorithm.

Principle 3 (a defined domain of applicability):

The domain of applicability of all the statistically significant QSPR models was analyzed in case of all two datasets using the standardization method. Thus the selection of the best QSPR model was done in corroboration with this principle.

Principle 4 (appropriate measures of goodness-of-fit, robustness and predictivity):

All the developed models were rigorously validated using internal, external and overall validation techniques. The quality of fitness and the predictive potential of the developed models were assessed based on the different validation metrics while the robustness of the models was judged using the randomization approach. The selection of the most significant models based on the acceptable values of the various validation metrics account for the compliance of the models with the 4th guideline

Principle 5 (a mechanistic interpretation):

All the descriptors appearing in the developed QSAR models could aptly define the essential structural attributes of the molecules imparting optimum endpoint values thus signifying suitable mechanistic interpretation of the developed models.

3.2.7. Software packages employed in the study

All the chemical structures were drawn using Marvin sketch (version 15.12.7.0), (<http://www.chemaxon.com>) software. Molecular descriptors were calculated using Dragon version 7 [81] and PaDEL Descriptor [82] software. Data pretreatment [117] was done by using Data Pretreatment GUI 1.2 software [116]. Data division was done by using Modified K-Medoid GUI 1.2 and Dataset Division GUI 1.2 [117] softwares. Stepwise regression was performed by using MINITAB Software (version 14.13) [72]. Other methods such as Partial Least Square, randomization, determination of applicability domain, best subset selection etc. was performed by software developed in our laboratory [89].

3.3. Study wise specific description of methodologies utilized in each study

3.3.1. Study 1: Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)

3.3.1.1. Selection of dataset

In this present work, we have developed QSPR models for data sets containing diverse organic contaminants with multi-end points of carbon nanotubes reported in the literature [118]. The dataset defined adsorption affinity properties (k_{∞}) of 59 organic contaminants by multi-walled carbon nanotubes (MWCNTs). All the endpoint values were taken in logarithmic scale for the modeling purposes. The data sets mainly contain adsorption data for synthetic organic compounds like pyrene, naphthalene, phenol, benzene, aniline, benzoate, chloroanisole, alcohol, acetophenone, isophoron, phenanthrene dicamba, atrazine, carbamazepine, pyrimidinone, acetamide, piperidine, propionitrile, acrylic acid, thiodiethanol, ethanol amine, cyclopentanone, acetone and ethylene glycol derivatives. K_{∞} is adsorption coefficients which could be obtained from isotherm data. K_{∞} is the ratio of q_e and C_e (q_e and C_e are solid and liquid phase equilibrium concentration respectively, at infinite dilution conditions with an average of 0.2% aqueous solubility).

3.3.1.2. Calculation of descriptors

The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiments.” All the datasets compounds were drawn by using Marvin Sketch software (<http://www.chemaxon.com>). The descriptors were calculated using two software tools namely Dragon software version 6 (http://www.taletc.mi.it/products/dragon_description.htm) and PaDel-descriptor (<http://www.yapcwsoft.com/dd/padeldescriptor>) software. In this work, we have calculated only a set of 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, atom type E-states, 2D atom pairs, molecular properties (using dragon software version 6) and ETA indices (using PaDel descriptor software).

3.3.1.3. Division of data into training and test set

Division of dataset is a very important step for QSAR. The present work deals with datasets containing diverse organic pollutants. The dataset compounds were divided into a training set and a test set using “Modified k-Medoid” clustering technique developed in our laboratory (https://teqip.jdvu.ac.in/QSAR_Tools/DTCLab). The clustering technique categorizes a set of compounds into cluster so that the compounds present in same cluster are similar to each other but when two compounds belong from two different clusters, are dissimilar in nature. The significant compounds within a cluster are called Medoid. Three clusters were generated for the dataset containing 59 organic pollutants. We have selected approximately 25% compounds of the total data set for test set and remaining 75% compounds selected for training set. The purpose of training set was to develop the model and test set was used to validate the model for calculation of different validation parameters.

3.3.1.4. Model development

After the very next step of dataset division, we have performed data pretreatment to remove inter correlated descriptors from the datasets. Prior to development of final models, we have tried to extract the important descriptors from the large pool of initial descriptors using

various variable selection strategies [119]. In this case, we have run stepwise regression and selected some descriptors. After removing the selected descriptors obtained from first stepwise regression run, we have run again stepwise regression using remaining pool of descriptors and we have repeated the same procedure. In this way, we have selected some manageable number of descriptors and make a pool (reduced pool of descriptors). After that we have run best subset selection for the dataset using reduced pool of descriptors developed in our laboratory [89]. Five (selected three models) and four (selected two models) descriptors models were generated in case of the dataset containing 59 organic pollutants. Among the equations generated from the best subset selection, we have selected five models for 59 organic pollutants based on MAE based criteria [104]. Finally, the selected models were run using intelligent consensus predictor (ICP) tool developed in our laboratory [72] to explore whether the quality of predictions of external compounds can be enhanced through an “intelligent” selection of multiple models (in this paper selected five models) as indicated by **Figure 2**

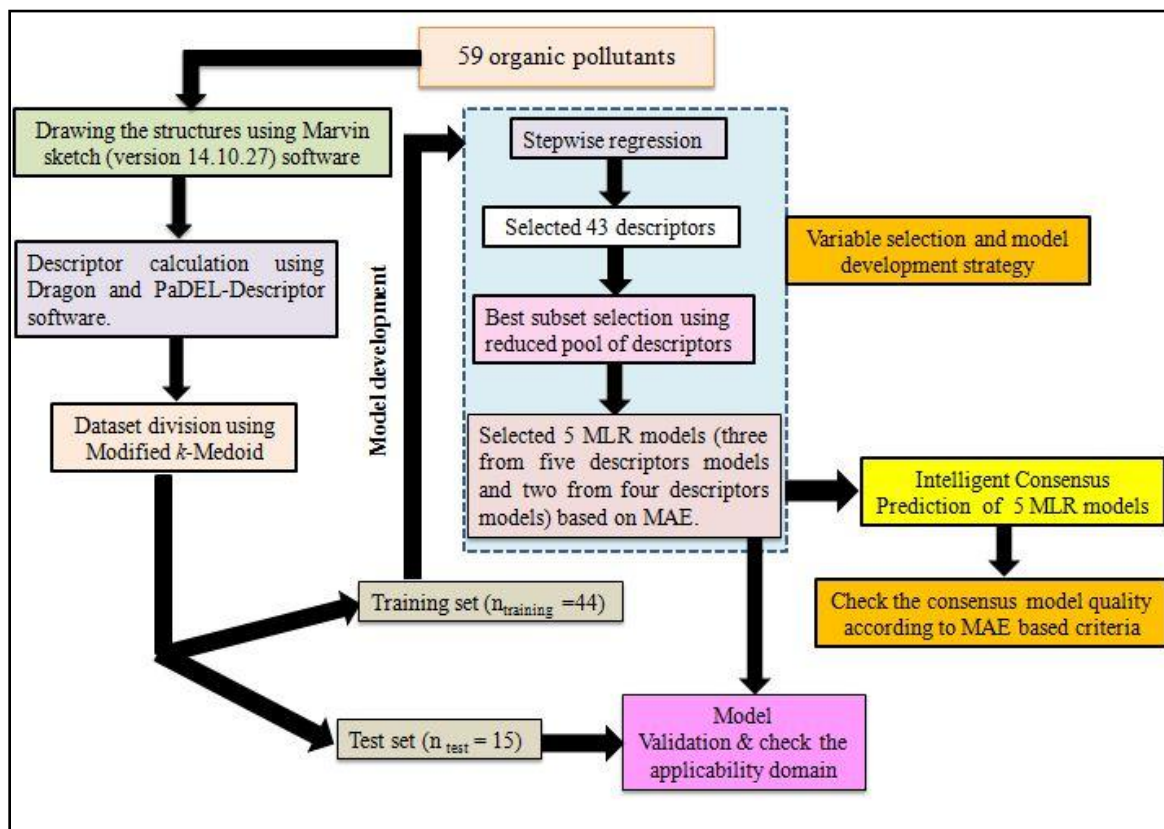


Figure 2. Schematic representation of the steps involved to develop the QSPR models of organic pollutants

3.3.2. Study 2: Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors

3.3.2.1. Selection of dataset

In the present work, we have developed nano-QSTR (Nano-Quantitative Structural Toxicity Relationship) models for 34 TiO₂ [79] NPs (nanoparticles) modified with varying amount and types of mixture of noble metals like Ag, Au and Pt (expressed in mole %). The cytotoxicity data towards the Chinese hamster ovary cell line (CHO-K1, ATCC® CCL-61™) was expressed in $-\log EC_{50}$ (negative logarithm of EC₅₀) for the development of Nano-QSTR models. For the purpose of modeling, all the nanoparticles were utilized and no single NPs were omitted. All the NPs used in the Nano-QSTR modeling were obtained from micro emulsion method [79].

3.3.2.2. Descriptor calculation

The Nano-QSTR models were developed from the fundamental information of the noble metal obtained from the periodic table to relate to the toxicity towards hamster ovary cell line and to investigate the modified TiO₂ based NPs in order to identify the key structural features responsible for the toxicity. For this purpose, 32 descriptors were taken from the periodic table, and one descriptor was derived from the primary descriptors. The periodic table descriptors can be adopted for calculation of integral additive descriptors of modified TiO₂ based NPs. To understand the structural changes in the heterogeneous NPs after modification of TiO₂ NPs using single metal clusters or with varying amount and types of noble metals with different concentration, the modified form of descriptor of equation (ii) is being used here for model development purpose. The calculation of the proposed mixture descriptors used in this work can be represented by the following **Eq. 3.29**,

$$D_{\text{mix}} = \% \text{ mol}_{\text{Me1}} \times P_1 + \dots + \% \text{ mol}_{\text{Men}} \times P_n \quad (3.29)$$

Here, D_{mix} means mixture descriptor, $\% \text{ mol}_{\text{Me1}}$ means concentration of each metal/component in the mixture – contribution by weight of metal in the NP sample of the mixture and P_n means the periodic table descriptor of individual metal. This method is used in order to treat each individual metal in the cluster as mixture system and each metal is described as a set of descriptors (calculated from the periodic table). The descriptor calculated for the metal (used

for coating of TiO₂ based NPs) from the periodic table is multiplied with the amount of the metal present in the mixture (TiO₂ based modified NPs), and the resulting sum defines a new set of descriptors of the complex mixture system. Periodic table based descriptors are more advantageous than others as they are easy to obtain without any significant calculation unlike quantum chemical descriptors [58, 59, 118].

3.3.2.3. Data set division

Data set division is a very crucial step for development of Nano-QSTR models. The data sets in the present work were divided into training (75% of the total dataset compounds) and test (25% of the total dataset compounds) sets based on “Modified k-Medoid” clustering technique. Clustering categorizes the compounds into clusters so that the compounds in the same clusters are similar and compounds from different clusters are dissimilar. This method tends to select the “k” most centered compounds or objects as the initial Medoid. Here, three clusters were obtained for 34 NPs. The purpose of the training set is to develop the models, and the test set is utilized for validating the models in terms of significance and robustness.

3.3.2.4. Model Development and Model selection

After performing the data set division, we have developed mono parametric Nano-QSTR models employing the Best Subset Selection v2.1 software [90] using mixed descriptors (33 descriptors) and selected five mono parametric LR models based on the MAE based criteria. After that, we have performed PLS regression of all the descriptors obtained from the previously developed five linear regression (LR) models with one latent variable.

The selection of the best descriptors for mono-parametric models is performed using software developed in our laboratory [89] Evaluation of the additive descriptors of mixture system is done along with the evaluation of different statistical parameters (R^2 , Q^2 , Q^2_{F1} , Q^2_{F2}) [121]. In the present work, we performed the Best Subset Selection v2.1 with the 34 periodic table descriptors and selected five LR models based on the MAE based criteria [104] after performing PLS by using 5 descriptors from 5LR models the applicability domain was checked. The detail of the method is presented in **Figure 3**.

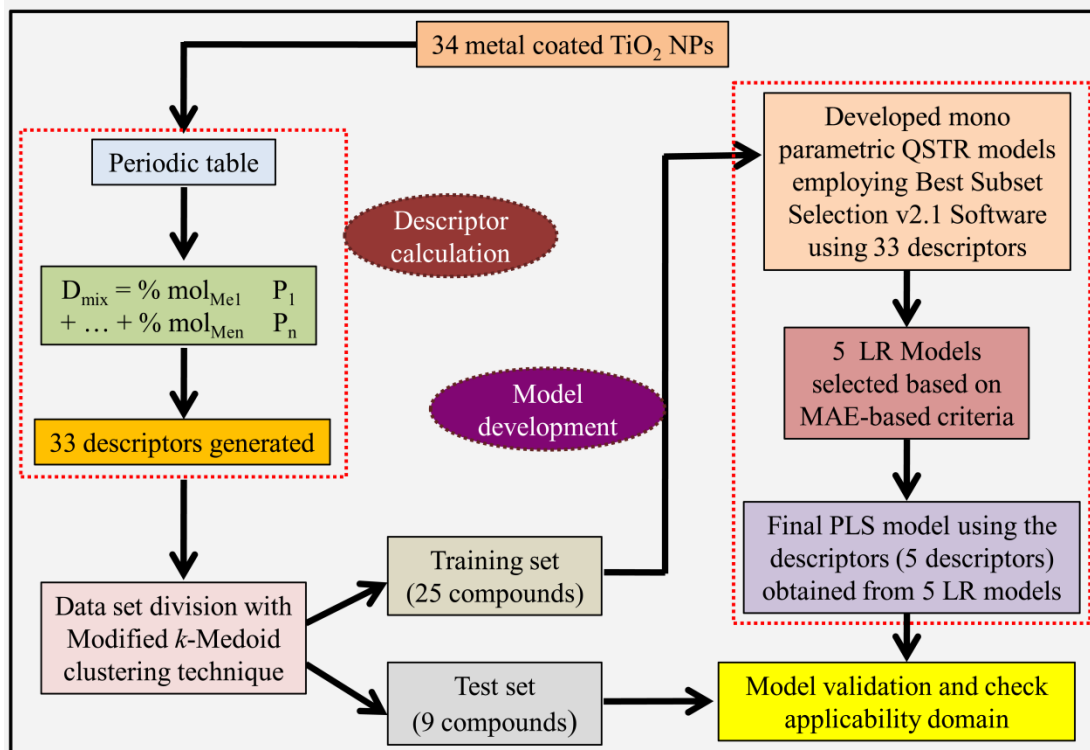


Figure 3. Schematic representation of the steps involved for the development of QSTR model

CHAPTER - 4

Results and discussions

4. RESULTS AND DISCUSSION

4.1. Study 1: Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)

We have developed QSPR models (five MLR models each for the dataset containing 59 (organic chemicals) with defined adsorption properties by MWCNTs using reduced descriptors pools obtained by different strategies as discussed in Materials and methods section. The calculated predicted values of all 5 MLR models are listed in **Table 4.1**. We have checked the statistical quality of all the individual models using both internal and external validation parameters which showed that the models are statistically significant (**Table 4.2**). We have checked MAE based criteria for all the models [104] The MAE based criteria of all the models are acceptable [104]. Besides the routinely used validation parameters, we have also checked the consensus prediction of the developed MLR models using a newly developed “Intelligent consensus predictor” tool [73] to check whether the quality of predictions of the test set compounds can be enhanced through an intelligent” selection of multiple MLR models. We found that the consensus predictions of multiple MLR models are better (based on MAE based criteria) than the results obtained from the individual model as shown in **Table 4.2** (here, the winner model is **CM3**). All the individual models are mentioned below and discussed the descriptors elaborately, in which n_{training} is the number of compounds used to develop the models and n_{test} is the number of compounds used for external prediction of the developed models. The leave-one-out (LOO) cross-validated correlation coefficient (Q^2) ($Q^2=0.863-0.895$) above the critical value of 0.5 signify the statistical reliability of the models. The predictability of the models was judged by means of predictive R^2 (R^2_{pred}) or Q^2_{F1} ($Q^2_{F1}=0.887-0.919$) and Q^2_{F2} ($Q^2_{F2}=0.886-0.919$) which show good predictive ability of the models. The statistical results of the model are summarized in **Table 4.2**.

The significant descriptors obtained from the five MLR models using the adsorption properties ($\log K_{oc}$) of 59 organic chemicals on MWCNTs are X0v, nArOH, B01[C-O], B06[C-Cl], Ui, F03 [O-O], F04 [N-O], ETA_BetaP, minsCH3, B03 [O-O] and nHBint4, which are regulating the adsorption property of the organic pollutants by MWCNTs. The

contribution of the descriptors can be easily identified from the regression coefficient of the independent variables. In this case, all the descriptors contributed positively (positive regression coefficient) except B01[C-O] descriptor (negative regression coefficient) as shown in **Table 4.3**. The significant descriptors obtained from the five MLR models (see in Box 1) for the adsorption properties ($\log K_{oc}$) of 59 organic chemicals on MWCNTs are X0v, nArOH, B01[C-O], B06[C-Cl], Ui, F03[O-O], F04[N-O], ETA_BetaP, minsCH₃, B03[O-O] and nHBint4, which regulate the adsorption property of the organic pollutants. The contribution of the descriptors can be easily identified from the regression coefficient of the independent variables. In this case, all the descriptors contributed positively (positive regression coefficient) except B01[C-O] descriptor (negative regression coefficient). The definition, contribution and frequency of the contributed descriptors are shown in **Table 4.3**. We have checked the applicability domain of the developed MLR models using standardization approach to confirm whether there is any compound present outside the applicability domain or not. It was found that one compound (compound number **41**) for model M1 is situated outside the applicability domain while compound number **56** is situated outside the domain of applicability in case of models M2, M3, M4 and M5. Though, these compounds showed good predictivity based on the models. The scatter plot of observed vs predicted adsorption coefficient for all the MLR models are shown in **Figure 4**.

Table 4.1: Observed and Predicted $\log K_{oc}$ values of 59 organic pollutants obtained from different MLR models.

Sl No.	Expt. ($\log K_{oc}$)	Predicted $\log K_{oc}$				
		Model 1	Model 2	Model 3	Model 4	Model 5
1	4.01	3.752	3.646	3.662	3.669	3.593
2*	1.63	1.234	1.071	1.081	1.131	1.09
3*	0.76	1.46	1.278	1.285	1.326	1.291
4	2.05	2.063	1.967	1.975	2.009	1.961
5	1.63	2.406	2.255	2.26	2.285	2.242
6	-0.45	-0.798	-0.889	-0.883	-0.783	-0.826
7*	-0.33	-0.054	-0.091	-0.09	-0.031	-0.054

8	1.17	1.3	1.45	1.443	1.46	1.449
9	0.33	0.557	0.434	0.449	0.526	0.468
10	2.38	2.224	2.285	2.31	2.325	2.269
11	-0.54	-0.504	-0.652	-0.652	-0.572	-0.602
12	0.21	0.355	0.262	0.066	0.297	-0.448
13	1.18	1.155	1.077	1.324	1.005	0.65
14	1.43	1.547	1.728	1.718	1.729	1.715
15	1.03	1.127	1.203	1.223	1.248	1.208
16	0.77	0.754	0.601	0.611	0.677	0.627
17*	-0.77	-0.411	-0.539	-0.537	-0.469	-0.491
18	-0.66	0.353	0.271	0.268	0.315	0.296
19	1.6	0.784	1.329	1.01	0.737	1.248
20*	0.72	0.852	0.712	0.723	0.781	0.74
21*	0.95	0.852	0.712	0.723	0.781	0.74
22	0.06	0.086	0.118	0.123	0.143	0.136
23	0.08	0.175	0.123	0.12	0.172	0.151
24	0.74	0.14	0.087	0.082	0.138	0.114
25	0.96	0.837	0.897	0.89	0.922	0.905
26	0.56	0.78	0.615	1.164	0.69	1.324
27	0.92	0.735	1.055	0.601	0.667	0.617
28	1.46	1.68	1.537	1.556	1.618	1.549
29	0.19	0.323	0.103	0.097	0.096	0.112
30*	0.26	0.451	0.268	0.262	0.259	0.277
31	0.5	0.474	0.459	0.454	0.498	0.478
32	0.76	0.765	0.361	0.345	0.318	0.356
33	0.82	0.506	0.676	0.681	0.687	0.68
34	0.04	0.363	0.47	0.494	0.592	0.513
35*	-0.32	-0.312	-0.489	-0.491	-0.428	-0.445
36	-0.9	-0.614	-0.633	-0.657	-0.615	-0.599
37	0.88	0.837	0.921	0.916	0.949	0.93

38	0.26	0.075	0.528	0.53	0.547	0.538
39	0.08	0.084	0.106	0.11	0.131	0.124
40*	0.7	0.361	0.682	0.677	0.684	0.689
41*	1.07	0.851	0.984	0.986	0.992	0.987
42	-0.46	-0.167	-0.366	-0.407	-0.388	-0.344
43	-0.15	-0.051	0.238	0.223	0.209	0.239
44	0.62	0.533	0.372	0.365	0.358	0.374
45	0.49	0.688	0.545	0.537	0.528	0.544
46	1.14	0.767	1.087	1.077	1.056	1.073
47	1.12	0.925	1.189	1.174	1.155	1.174
48	0.62	0.147	0.094	0.505	0.144	0.729
49*	1.44	1.123	1.193	1.211	1.232	1.202
50	1.28	1.624	1.334	1.335	1.331	1.325
51	0.79	0.696	0.679	0.671	0.705	0.691
52	1.89	1.818	1.947	1.967	1.967	1.93
53	0.56	0.627	0.677	0.673	0.713	0.691
54	0.65	0.618	0.673	0.668	0.709	0.687
55*	0.51	0.623	0.669	0.666	0.704	0.687
56*	0.01	0.514	0.142	0.082	0.002	0.114
57	2.73	2.386	1.767	1.772	1.807	1.763
58	2.72	2.809	2.94	2.957	2.986	2.915
59*	3.29	3.007	2.871	2.882	2.898	2.853

*indicates test set compounds

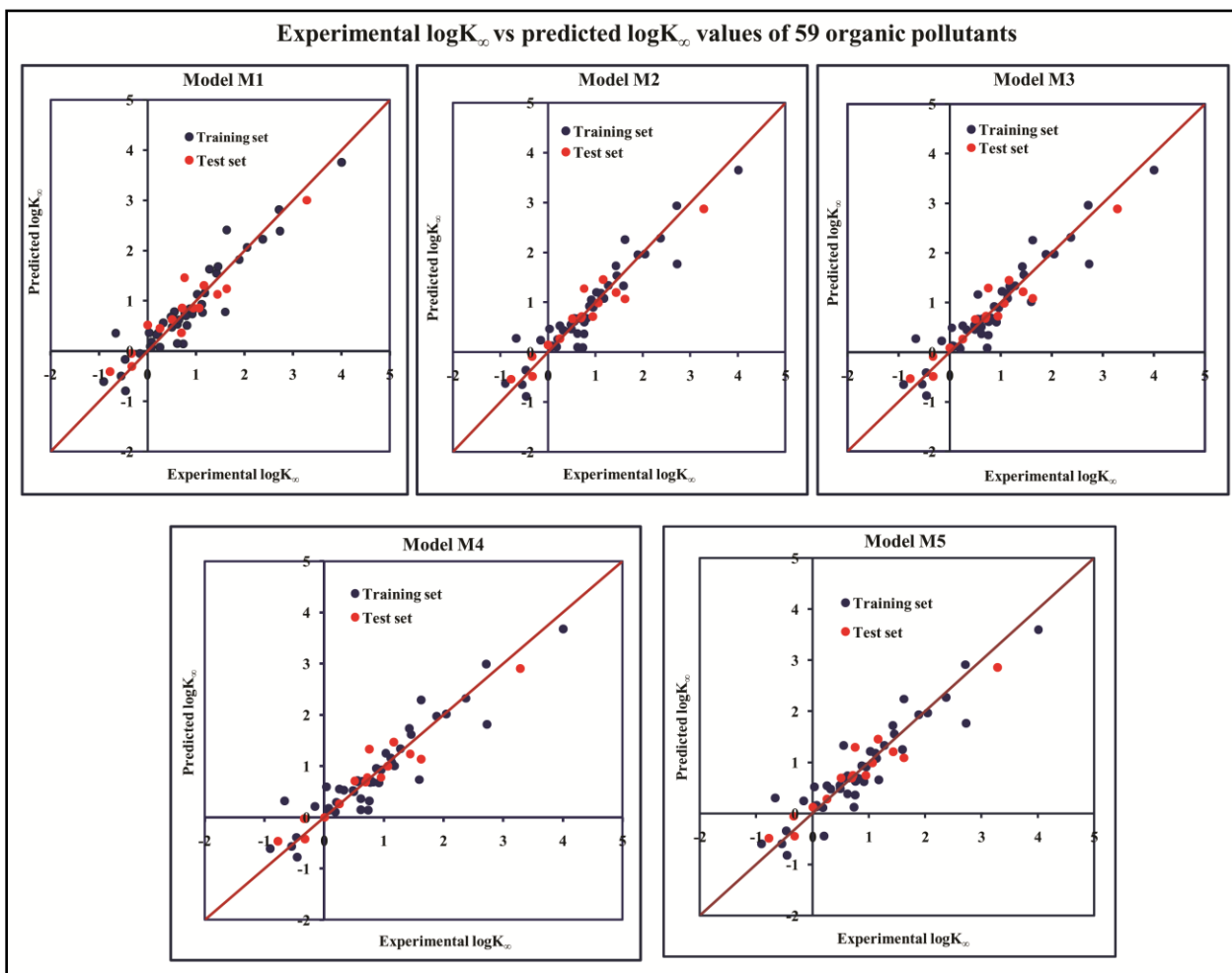


Figure 4. The scatter plot of the observed and the predicted adsorption coefficient property ($\log K_{oc}$) of the developed MLR models (models M1-M5).

Table 4.2: Calculated Statistical parameters of individual and Consensus models.

Dataset	Type of model		Training set statistics					Test set statistics							
			Model R ²	Model Q ² _(LOO)	MAE _{train}	$\overline{r}_{m(LOO)}^2$	$\Delta r_{m(LOO)}^2$	R ² _{pred} or Q ² _{F1}	Q ² _{F2}	CCC	$\overline{r}_{m(test)}^2$	$\Delta r_{m(test)}^2$	MAE (100%)	MAE (95%)	MAE
59 organic contaminants	Individual Models	IM1	0.920	0.895	Good	0.8512	0.0777	0.887	0.886	0.934	0.745	0.104	0.271	0.240	Good
		IM2	0.912	0.892	Good	0.8481	0.0790	0.916	0.915	0.952	0.817	0.072	0.221	0.197	Good
		IM3	0.905	0.880	Good	0.8321	0.0751	0.919	0.919	0.954	0.825	0.069	0.213	0.189	Good
		IM4	0.893	0.872	Good	0.8206	0.0920	0.918	0.917	0.953	0.806	0.074	0.213	0.187	Good
		IM5	0.893	0.863	Good	0.8083	0.0864	0.915	0.914	0.950	0.798	0.076	0.222	0.199	Good
	Consensus Models	CM0	-	-	-	-	-	0.917	0.916	0.952	0.800	0.074	0.227	0.203	Good
		CM1	-	-	-	-	-	0.917	0.916	0.952	0.800	0.074	0.227	0.203	Good
		CM2	-	-	-	-	-	0.919	0.919	0.953	0.803	0.073	0.221	0.196	Good
		CM3	-	-	-	-	-	0.935	0.935	0.962	0.812	0.059	0.187	0.163	Good

CM0= “ordinary” consensus predictions.

CM1 = Average of predictions from individual models IM1 through IM5

CM2 = Weighted average predictions from individual models IM1 through IM5

CM3 = Best selection of predictions (compound-wise) from individual models IM1 through IM5

***Note that we have run the “Intelligent consensus predictor tool” using AD: No; Dixon Q-test: No; Euclidean distance: No.**

Table 4.3: Definition, frequency and contribution of all the descriptors obtained from the MLR models (models developed by using 59 organic pollutants).

Sl. no.	Name of descriptors	Contribution	Discussion	Mechanism	Frequency of descriptors
1	X0v	+ve	Valence connectivity index order 0.	Hydrophobic interaction	5
2	nArOH	+ve	Number of aromatic hydroxyls.	Hydrogen bonding interactions.	1
3	B01[C-O]	-ve	Presence/absence of C-O fragment at topological distance 1.		1
4	B06[C-Cl]	+ve	Presence/absence of C-Cl fragment at topological distance 6.	Here, size plays an important role for adsorption affinity. The size enhances the surface area of molecules which regulate the hydrophobic interactions between organic pollutants and MWCNTs.	1
5	Ui	+ve	Unsaturation index.	π - π interactions.	1
6	F03[O-O]	+ve	Frequency of O-O at topological distance 3.	Enhanced electron density, electrostatic interactions.	2
7	F04[N-O]	+ve	Frequency of N-O at topological distance 4.	Hydrogen bonding and electrostatic interactions.	1
8	ETA_BetaP	+ve	A measure of electronic features of the molecules relative to molecular size.	Electrostatic interactions.	4

9	minsCH ₃	+ve	Minimum atom E-state -CH ₃ .	Hydrophobic interactions.	4
10	B03[O-O]	+ve	Presence/absence of O-O at topological distance 3.	Electrostatic interactions, hydrogen bonding.	1
11	nHBint4	+ve	Count of E-state descriptors of strength for potential hydrogen bond of path length 4.	Hydrogen bonding interaction.	2

Box 1

Model M1.

$$\log k_{\infty} = -4.62(\pm 0.337) + 0.834(\pm 0.155) \times Ui + 0.663(\pm 0.220) \times B06[C - Cl] \\ + 0.641(\pm 0.057) \times X0v + 0.600(\pm 0.091) \times nArOH - 0.611(\pm 0.121) \times B01[C - O]$$

$$n_{training} = 44, R^2 = 0.920, R_{adj}^2 = 0.908, S = 0.294, F = 85.93, PRESS = 4.267,$$

$$Q^2 = 0.895, \overline{r_{m(LOO)}^2} = 0.851, \Delta r_{m(LOO)}^2 = 0.078, MAE = good,$$

$$n_{test} = 15, Q_{F1}^2 = 0.887, Q_{F2}^2 = 0.886, \overline{r_{m(test)}^2} = 0.745, \Delta r_{m(test)}^2 = 0.104, CCC = 0.934, MAE = good$$

Model M2.

$$\log k_{\infty} = -8.51(\pm 0.722) + 0.803(\pm 0.048) \times X0v + 0.681(\pm 0.146) \times F03[O - O] \\ + 0.415(\pm 0.144) \times F04[N - O] + 3.27(\pm 0.491) \times ETA_BetaP + 0.204(\pm 0.067) \times \min sCH_3$$

$$n_{training} = 44, R^2 = 0.912, R_{adj}^2 = 0.900, S = 0.306, F = 78.66, PRESS = 4.356,$$

$$Q^2 = 0.892, \overline{r_{m(LOO)}^2} = 0.848, \Delta r_{m(LOO)}^2 = 0.079, MAE = good,$$

$$n_{test} = 15, R_{pred}^2 = 0.916, Q_{F2}^2 = 0.915, \overline{r_{m(test)}^2} = 0.817, \Delta r_{m(test)}^2 = 0.072, CCC = 0.952, MAE = good$$

Model M3.

$$\log k_{\infty} = -8.68(\pm 0.746) + 0.802(\pm 0.050) \times X0v + 0.603(\pm 0.272) \times B03[O - O] \\ + 3.39(\pm 0.503) \times ETA_BetaP + 0.213(\pm 0.069) \times \min sCH_3 + 0.412(\pm 0.148) \times nHBint_4$$

$$n_{training} = 44, R^2 = 0.905, R_{adj}^2 = 0.893, s = 0.318, F = 72.57, PRESS = 4.840,$$

$$Q^2 = 0.880, \overline{r_{m(LOO)}^2} = 0.832, \Delta r_{m(LOO)}^2 = 0.075, MAE = good,$$

$$n_{test} = 15, Q_{F1}^2 = 0.919, Q_{F2}^2 = 0.919, \overline{r_{m(test)}^2} = 0.825, \Delta r_{m(test)}^2 = 0.069, CCC = 0.954, MAE = good.$$

Model M4.

$$\log k_{\infty} = -8.72(\pm 0.782) + 0.785(\pm 0.052) \times X0v + 0.650(\pm 0.158) \times F03[O - O] \\ + 3.51(\pm 0.527) \times ETA_BetaP + 0.202(\pm 0.073) \times \min sCH_3$$

$$n_{training} = 44, R^2 = 0.893, R_{adj}^2 = 0.882, s = 0.334, F = 81.11, PRESS = 5.164,$$

$$Q^2 = 0.872, \overline{r_{m(LOO)}^2} = 0.821, \Delta r_{m(LOO)}^2 = 0.092, MAE = good$$

$$n_{test} = 15, Q_{F1}^2 = 0.918, Q_{F2}^2 = 0.917, \overline{r_{m(test)}^2} = 0.806, \Delta r_{m(test)}^2 = 0.074, CCC = 0.953, MAE = good$$

Model M5.

$$\log k_{\infty} = -8.42(\pm 0.773) + 0.785(\pm 0.052) \times X0v + 3.29(\pm 0.526) \times ETA_BetaP \\ + 0.199(\pm 0.072) \times \min sCH_3 + 0.566(\pm 0.137) \times nHBint_4$$

$$n_{training} = 44, R^2 = 0.893, R_{adj}^2 = 0.882, s = 0.333, F = 81.33, PRESS = 5.543,$$

$$Q^2 = 0.863, \overline{r_{m(LOO)}^2} = 0.808, \Delta r_{m(LOO)}^2 = 0.086, MAE = good$$

$$n_{test} = 15, Q_{F1}^2 = 0.915, Q_{F2}^2 = 0.914, \overline{r_{m(test)}^2} = 0.798, \Delta r_{m(test)}^2 = 0.076, CCC = 0.950, MAE = good.$$

4.1.1. The descriptors related to hydrogen bonding interactions

The functional group count descriptor, n_{ArOH} , represents number of aromatic hydroxyl group present in the compound. This descriptor influences the adsorption property of organic pollutants by MWCNTs as indicated by its positive regression coefficient. Thus, the compounds containing large number of aromatic hydroxyl group may enhance the adsorption property of organic pollutants by MWCNTs as shown in compounds **13 (pyrogallol)** (containing 3-OH groups), **5 (2-phenyl phenol)** (containing 1-OH group) and **14 (2,4,6 trichloro phenol)** (containing 1-OH group), whereas the compounds containing no aromatic hydroxyl group are detrimental for the adsorption affinity of organic pollutants by MWCNTs as shown in compounds **18 (4-chloroaniline)**, **36 (benzyl alcohol)** and **42 (phenethyl alcohol)** (these compounds contain no aromatic hydroxyl group). Although some compounds containing no aromatic hydroxyl group still show high adsorption affinity of the organic pollutants by MWCNTs, it is due to some other dominating descriptors present in the model. Thus, the substitution of electron donating group like hydroxyl group in aromatic ring of organic pollutants could enhance the adsorption on MWCNTs.

A 2D atom pair descriptor, $F_{04}[\text{N-O}]$, indicates the frequency of N-O fragment at topological distance 4. The positive regression coefficient of the descriptor suggests that increase of N-O fragment at topological distance 4 is directly proportional to the adsorption affinity of organic pollutants. The higher number of fragment correlates to higher adsorption property as observed in case of compounds **19 (2-nitroaniline)** and **27 (3-nitrophenol)** while the absence of such fragment at topological distance 4 has no influence in the adsorption by MWCNTs as shown in compounds **18 (4-chloroaniline)**, **36 (benzyl alcohol)** and **42 (phenethylalcohol)**. This descriptor also indicated that the frequency of two electronegative atoms of organic pollutants (either electron donating or electron withdrawing group; in case of compound number **19**, nitrogen (-NH₂ group) acts as electron donor and oxygen (-NO₂ group) acts as a electron withdrawing group whereas in case of compound number **27**, nitrogen (-NO₂ group) acts as a electron withdrawing group and oxygen (-OH group) acts as a electron donating group) should be situated at topological distance 4 for better adsorption on MWCNTs.

An E-state descriptor, n_{HBint4} , a PaDel descriptor, indicates count of potential internal hydrogen bonds separated by four edges. The positive regression coefficient suggests that

propensity of hydrogen bond of organic pollutants has a dominating role to enhance the adsorption property. Thus, the organic pollutants bearing hydrogen bonded group separated by four path length are conducive for adsorption affinity as shown in compounds **13 (pyrogallol)**, **19 (2-nitroaniline)** and **48 (3-chlorophenol)** whereas absence of such fragment in organic pollutants are detrimental for the adsorption affinity as shown in compounds **6 (benzene)**, **11 (phenol)** and **42 (phenethyl alcohol)**.

B03[O-O] is a 2D atom pair descriptor that indicates presence or absence of O-O fragment at topological distance 3. The positive regression coefficient of the descriptor indicates that higher the frequency of this fragment, higher is the adsorption affinity. Thus, the presence of O-O fragment at topological distance 3 favors the adsorption of organic pollutants by MWCNTs as shown in compound nos. **12 (catechol)** and **13 (pyrogallol)** while on the contrary the compound nos. **6 (benzene)**, **42 (phenethyl alcohol)** and **36 (benzyl alcohol)** have less influence to adsorption property as these compounds having no such fragments at topological distance 3.

Hydrogen bonding is one of the key mechanisms for the adsorption of organic contaminants on CNTs. The information obtained from the descriptors nArOH, F04[N-O], nHBint4, F03[O-O] and B03[O-O] suggested that there may be some hydrogen bonding interactions between organic pollutants and MWCNTs which regulate the adsorption affinity (Fig. 3) of organic pollutants to MWCNTs. In case of the descriptor nArOH, the aromatic hydroxyl group may form hydrogen bonds with the hydroxyl/carboxylic groups on the CNTs surface and the hydrogen bonds may also form between the surface-adsorbed aromatic hydroxyl group containing organic pollutants (phenolics) and dissolved phenolics. Here, the hydroxyl group is always connected with an aromatic ring. Thus, it is obvious that this aromatic ring of organic pollutants itself can interact with CNTs by π - π interactions. The descriptor, F04[N-O] also suggested that besides the hydrogen bonding interaction, there may also be a chance to form electrostatic interaction. The electron withdrawing group like NO₂ may also strengthen the π - π interactions formed between the benzene derivatives (acting as π -acceptor) and CNTs (acting as π -donor). In case of B03[O-O], two oxygen atom (hydroxyl group) is separated by topological distance 3. Thus, these oxygen atoms substituted in a form of hydroxyl group may interact with CNTs by hydrogen bonding interactions. These two electronegative atoms of organic pollutants situated at the topological distance 3 may also

interact electrostatically with CNTs. This fragment may also strengthen the π - π interactions formed between organic pollutants and MWCNTs [122-123]. Note that, though C-O bond is detrimental for adsorption affinity of organic pollutants on CNTs as discussed later, but frequency of O-O fragment (O atom attached with any carbon atom i.e., C-O bond (one oxygen atom)); here, O-O fragment (containing two oxygen atoms) suppress the detrimental effect of C-O group) at topological distance 3 and is influential for the adsorption affinity of organic pollutants on MWCNTs. The descriptors involved for hydrogen bonding interaction between the organic pollutants and MWCNTs are depicted in **Figure 5**.

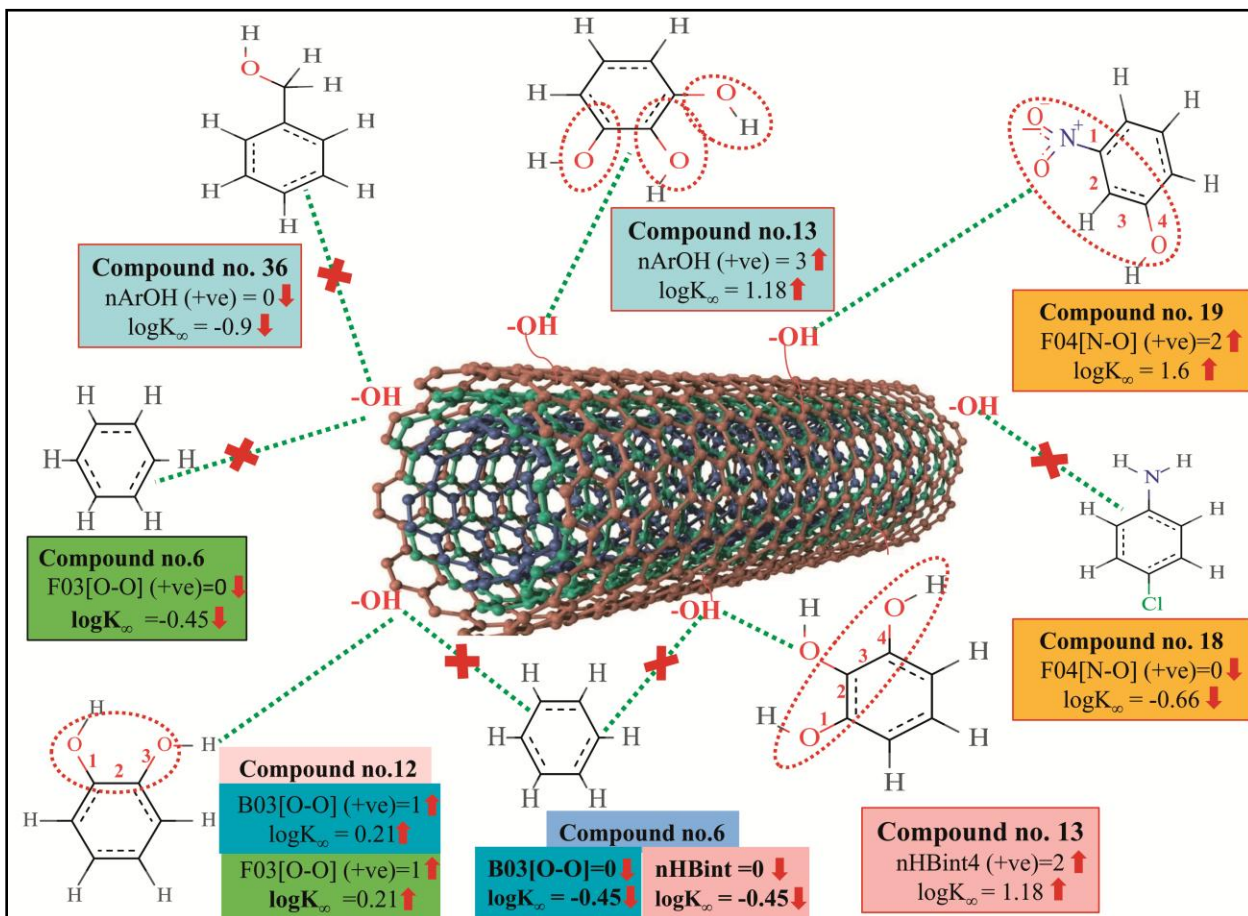


Figure 5. Mechanistic interpretation of the descriptors related to hydrogen bonding interaction between organic pollutants and MWCNTs.

4.1.2. The descriptors related to hydrophobic interactions

A 2D atom pair descriptor, B06[C-Cl], represents presence or absence of C-Cl bond at topological distance 6. The positive regression coefficient of this parameter suggested that

presence of such fragment at topological distance 6 enhances the adsorption affinity of organic pollutants towards the MWCNTs as shown in compounds **50** (**4-chloroacetophenone**) and **57** (**2-chloronaphthlene**). On the other hand, compounds like **11** (**phenol**), **22** (**4-methylphenol**) and **43** (**3-methylbenzyl alcohol**) show poor adsorption affinity by the MWCNTs due to absence of such fragment.

The descriptor $X0v$ indicates valence connectivity index of order 0, which can be calculated through Kier and Hall's connectivity index as shown below. This descriptor contributed positively towards the adsorption affinity of organic pollutants to the MWCNTs. Thus, size of the organic pollutants plays a crucial role to regulate the adsorption affinity of organic pollutants to MWCNTs. It has been found that on increasing the numerical value of this descriptor, the adsorption affinity of organic pollutants by MWCNTs also increases as shown in case of compounds **1** (**pyrene**), **58** (**azobenzene**), **5** (**2-phenyl phenol**) (bigger in size) while the adsorption affinity of organic pollutants by MWCNTs decreases in case of compounds **6** (**benzene**), **11** (**phenol**) and **36** (**benzyl alcohol**) (smaller in size).

Valence connectivity index of zeroth can be calculated by (4.1):

$$X0v = \sum_{i=1}^n (\delta_i^v)^{-0.5}$$

$$\delta_i^v = \frac{Z_i^v - hi}{Z_i - Z_i^v - 1} \quad (4.1)$$

In the above equation, δ_i^v =valence vertex degree, Z_i^v = valence electron in the i^{th} atom, hi =number of hydrogen atoms connected to the i^{th} atom, Z_i =number of all electron in the i^{th} atom.

The E-state indices of a particular atom in a certain molecule provide information on its electronic state of that particular atom which in turn depends upon π bonds, lone pair of electrons and δ bonds that inform the quantitative availability of valence electrons [124]. The descriptor minsCH_3 indicates minimum atom type E-state CH_3 . The positive regression coefficient of this descriptor indicates that presence of CH_3 group has an important role to influence the adsorption property of organic pollutants. The numerical value of this descriptor is directly proportional to the adsorption property, which suggests that with increasing the numerical value of this descriptor, the adsorption affinity of the organic pollutants also increases as evidenced by compounds **10** (**2,4-dinitrotoluene**), **50** (**4-**

chloroacetophenone) and **52 (1-methylnaphtalene)**. On the other hand, the adsorption affinity of organic pollutants decreases with absence of CH₃ group as showed in compounds **6 (benzene)**, **11 (phenol)** and **36 (benzyl alcohol)**.

Hydrophobic interaction between organic pollutants and CNTs is also an important mechanism for better adsorption. The descriptors, B06[C-Cl], X0v and minsCH₃ suggested that the organic pollutants may adsorb to the MWCNTs by hydrophobic interaction. In case of B06[C-Cl] and X0v, size of molecules (for B06[C-Cl], the distance between C and Cl atoms is six which reflects on size of the molecules) plays an important role for adsorption affinity. The size enhances the surface area of molecules which can regulate the hydrophobic interactions between organic pollutants and MWCNTs. The methyl group (information obtained from minsCH₃ descriptor) and CNTs are hydrophobic in nature. Thus, an increase in the minsCH₃ value would indicate a higher degree of unsaturation, hence would enhance the reactivity. So, there is a chance to form hydrophobic interactions between organic pollutants and MWCNTs which reflects better adsorption. The descriptors involved for hydrophobic interactions between organic pollutants and CNTs are depicted in **Figure 6**.

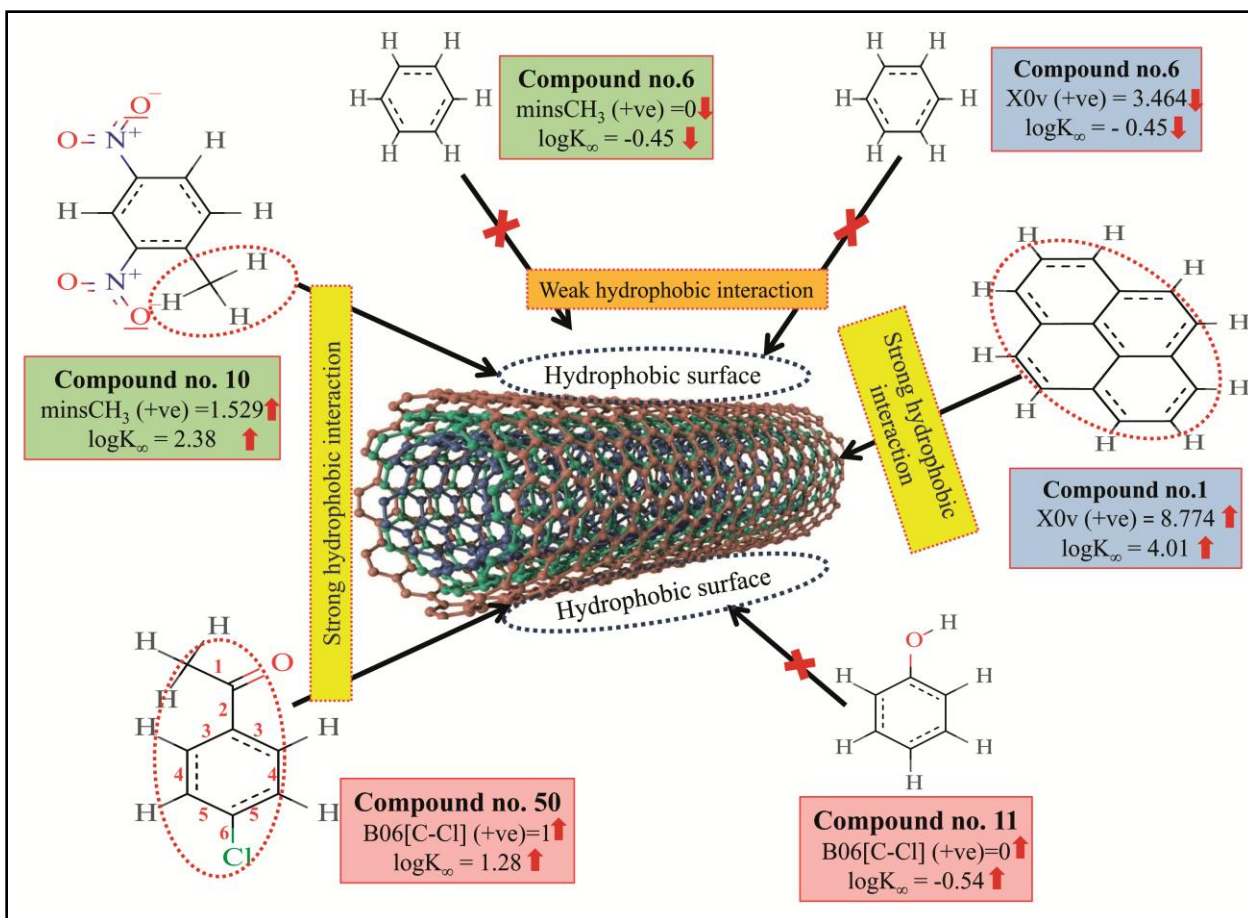


Figure 6. Mechanistic interpretation of the descriptors related to hydrophobic interaction between organic pollutants and MWCNTs.

4.1.3. The descriptors related to π - π interactions

The descriptor, U_i , gives information about the Unsaturation index, which contributed positively towards the adsorption affinity of organic pollutants by MWCNTs indicated by positive regression coefficient. From this descriptor, it has been suggested that presence of unsaturation in organic pollutants plays a crucial role to enhance the adsorption affinity as shown in compounds **1** (pyrene), **10** (2,4-dinitrotoluene) and **58** (azobenzene) (the numerical values of this descriptor are 3.392, 3 and 3 respectively) and vice versa in case of compounds **11** (phenol), **36** (benzyl alcohol) and **42** (phenethyl alcohol) (the numerical values of this descriptor are 2 in each compound). Here, the compounds, **1** (pyrene), **10** (2,4-dinitrotoluene) and **58** (azobenzene) have higher range of unsaturation index value due to presence of large number of double bonds.

An ETA index, ETA_BetaP gives a measure of sigma, pi and non-bonded (i.e., lone pair capable of forming resonance with aromatic system) electrons relative to the molecular size. Therefore, electron-richness (unsaturation) relative to the molecular size of organic pollutants is an important parameter to regulate the adsorption property. The positive regression coefficient of this parameter indicates that electron density of molecules should be higher for increasing the adsorption affinity of organic pollutants by MWCNTs as found in the compound no. **1 (pyrene)**, **28 (1,3-dinitrobenzene)** and **58 (azobenzene)** whereas the compounds with low electron density show lower range of adsorption affinity as shown in compound nos. **36 (benzyl alcohol)**, **42 (phenethyl alcohol)** and **43 (3-methylbenzyl alcohol)**. Thus, it can be concluded that the molecules should be electron-rich for higher adsorption property of organic pollutants.

The π - π interaction is another most important mechanism involved for the adsorption of organic pollutants to CNTs. The information obtained from U_i and ETA_BetaP descriptors suggested that the organic pollutants can adsorb to MWCNTs by strong π - π interaction. Whereas, the descriptors B03[O-O], F03[O-O] and F04[N-O] suggested that the [O-O] fragments at topological distance 3 and the [N-O] fragments at topological distance 4 may strengthen the π - π interaction formed between organic pollutants and MWCNTs. The descriptor, U_i suggested that unsaturation plays a crucial role for adsorption of organic pollutants to MWCNTs. CNTs also contains large number of double bonds (unsaturation). So, there is a chance to form strong π - π interactions between organic pollutants and MWCNTs which reflects better adsorption of these pollutants to MWCNTs. Hence, a higher number of double bonds of organic pollutants enhance the adsorption affinity to MWCNTs. The descriptor, ETA_BetaP suggested that unsaturation (electron-richness) relative to the molecular size of organic pollutants plays a crucial role to regulate the adsorption property. From this descriptor, it can be suggested that the adsorption affinity of organic pollutants to MWCNTs is increased due to the π - π interaction. The descriptors involved for π - π interaction between organic pollutants and CNTs are described graphically in **Figure 7**.

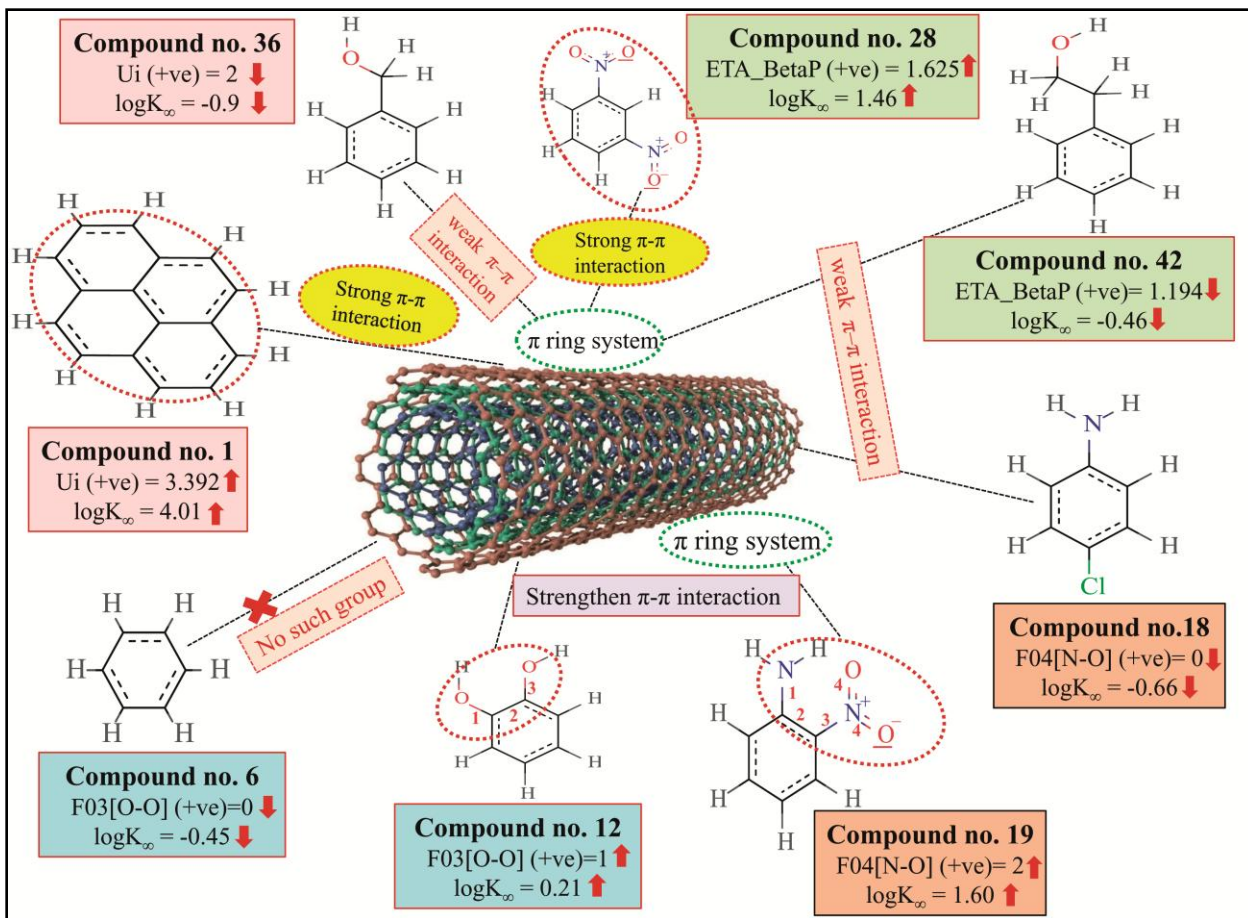


Figure 7. Mechanistic interpretation of the descriptors related to π - π interaction between organic pollutants and MWCNTs.

4.1.4. The descriptors related to electrostatic interactions

F03[O-O], a 2D atom pair descriptor, indicates frequency of O-O fragment at topological distance 3. The positive regression coefficient of this descriptor suggested that presence of higher number of O-O bonds at topological distance 3 might be beneficial for the adsorption affinity of organic pollutants by MWCNTs as shown in compounds **12** (catechol) and **13** (pyrogallol) whereas the opposite happens in case of compounds **6** (benzene), **42** (phenethyl alcohol) and **43** (3-methylbenzyl alcohol) (where, no O-O fragment present at topological distance 3). This fragment may also strengthen the π - π interactions formed between organic pollutants and MWCNTs [125-126]. Like B03[O-O], this descriptor also suppresses the detrimental effect of C-O group as discussed earlier in this section.

The information obtained from the descriptors, F03[O-O], B03[O-O] and F04[N-O] suggested that the organic pollutants can adhere to the surface of MWCNTs by strong electrostatic interaction. The descriptors F03[O-O] and B03[O-O] indicate that the frequency or presence/absence of two electronegative atoms (electron donating group) at the topological distance 3 is essential to enhance the adsorption affinity of organic pollutants to MWCNTs. Thus, there may be a chance to form electrostatic interactions between organic pollutants (negative charged atom like oxygen atom of hydroxyl group) and MWCNTs (sidewall of the CNTs are electrically polarizable thus polar molecules can easily adhere to their surface). The descriptors involved for electrostatic interaction between organic pollutants and CNTs are represented graphically in **Figure 8**.

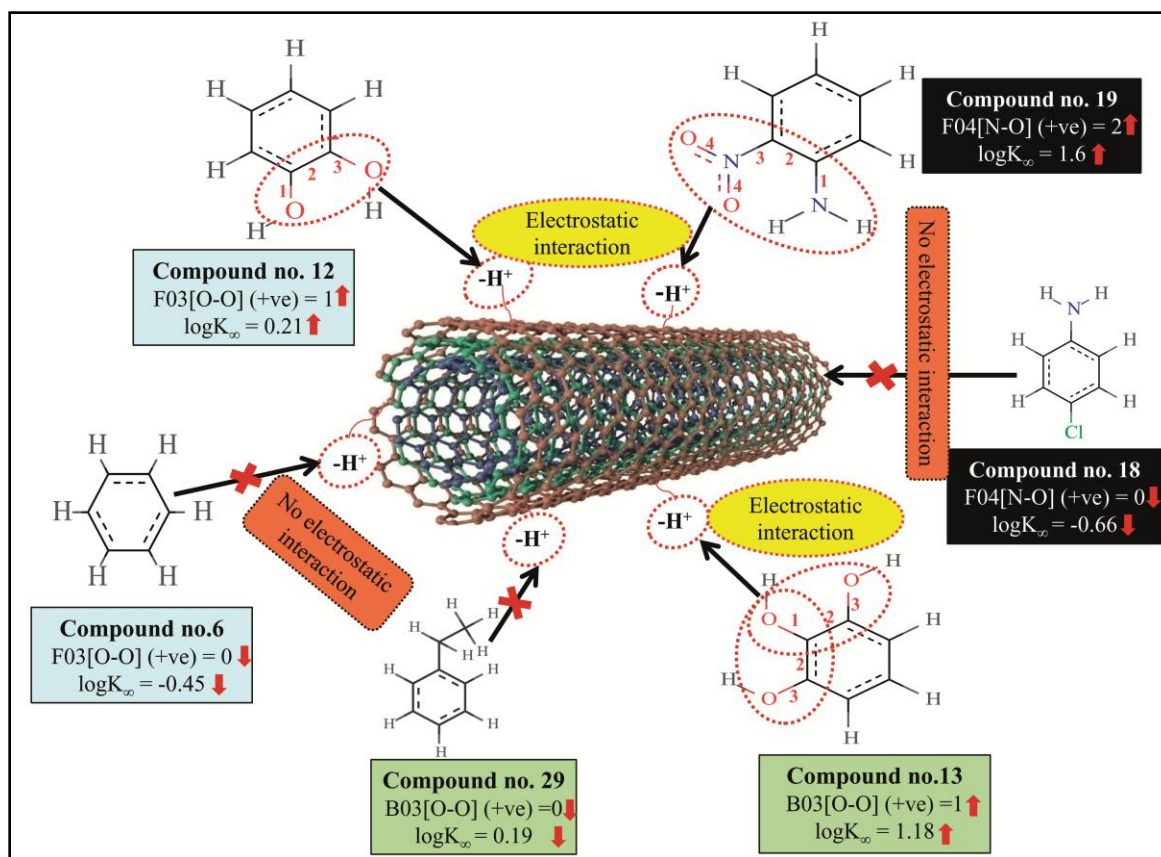


Figure 8. Mechanistic interpretation of the descriptors related to electrostatic interaction between organic pollutants and MWCNTs.

The 2D atom pair descriptor, B01[C-O], indicates the presence or absence of the C-O bond at topological distance 1. The negative regression coefficient of the descriptor supports that presence of this fragment at topological distance one is detrimental to the adsorption affinity

of organic pollutants by MWCNTs though it can form hydrogen bond with MWCNTs. As for example, compounds like **1** (pyrene), **57** (2-chloronaphthalene) and **58** (azobenzene) have higher adsorption affinity value due to absence of such fragments at topological distance 1 whereas compounds like **11** (phenol), **36** (benzyl alcohol) and **42** (phenethyl alcohol) have lower adsorption affinity due to presence of one C-O bonds each.

4.2. Study 2: Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors

Based on the cytotoxicity data of 34 TiO₂ modified NPs towards Chinese hamster ovary cells and easily calculated 34 periodic table based descriptors, we have developed five simple but statistically significant LR based Nano-QSTR models. We have checked both the internal ($R^2=0.922-0.926$; $Q^2=0.907-0.911$; $R^2_{adj}=0.918-0.922$) and external ($Q^2_{F1}=0.930-0.938$; $Q^2_{F2}=0.924-0.932$) validation parameters of all the individual models which showed good *in silico* predictivity of the models as depicted in **Table 4.3**. For the external validation, Q^2_{F1} or R^2_{pred} and Q^2_{F2} metrics were used, and their values are much higher than the threshold value, i.e., 0.5. MAE based criteria were also checked for the models and each individual model passed the MAE based criteria. We have calculated r_m^2 parameters like $\overline{r_{m(loo)}^2}$ and $\Delta r_{m(loo)}^2$ and $Q^2_{(Loo)}$ for the internal set and $r_{m(test)}^2$ and $\Delta r_{m(test)}^2$ for the external set, and the resultant values passed the critical values proving statistical reliability of the models. The descriptors obtained from the five LR models are discussed elaborately below. In equation, $n_{training}$ means number of compounds in the training set/internal set used to develop the models, and n_{test} means number of compounds in the test set/external set used to judge the quality of the models. The test compounds showed good predictivity based on the models. The scatter plot of observed vs predicted cytotoxicity for all the LR and PLS models are shown in **Figure 9**.

Table 4.3: Statistical quality and validation parameters obtained from the developed LR and PLS models.

Types of models		Descriptors	Training set statistics					Test set statistics						
			Model R^2	Model $Q^2_{(LOO)}$	MAE _(95%)	$\overline{r^2_{m(LOO)}}$	$\Delta r^2_{m(LOO)}$	R^2_{pred} or Q^2_{F1}	Q^2_{F2}	CCC	$\overline{r^2_{m(test)}}$	$\Delta r^2_{m(test)}$	MAE _(95%)	MAE
LR	M1	Electrochemical Equivalent (g/amp-hr)	0.926	0.911	Moderate	0.880	0.045	0.937	0.931	0.959	0.828	0.049	0.066	Good
	M2	2nd ionization (kJ/mol)	0.923	0.909	Moderate	0.877	0.047	0.938	0.932	0.960	0.843	0.042	0.057	Good
	M3	Amount of Ag	0.922	0.907	Moderate	0.875	0.048	0.930	0.924	0.956	0.806	0.059	0.073	Good
	M4	Thermal conductivity (W/(m. k))	0.922	0.907	Moderate	0.874	0.048	0.934	0.928	0.958	0.829	0.052	0.077	Good
	M5	covalent radius (pm)	0.923	0.909	Moderate	0.877	0.047	0.938	0.932	0.960	0.842	0.042	0.056	Good
PLS	P1	All 5 descriptor with one latent variable	0.925	0.911	Moderate	0.883	0.0483	0.944	0.938	0.969	0.922	0.031	0.068	Good

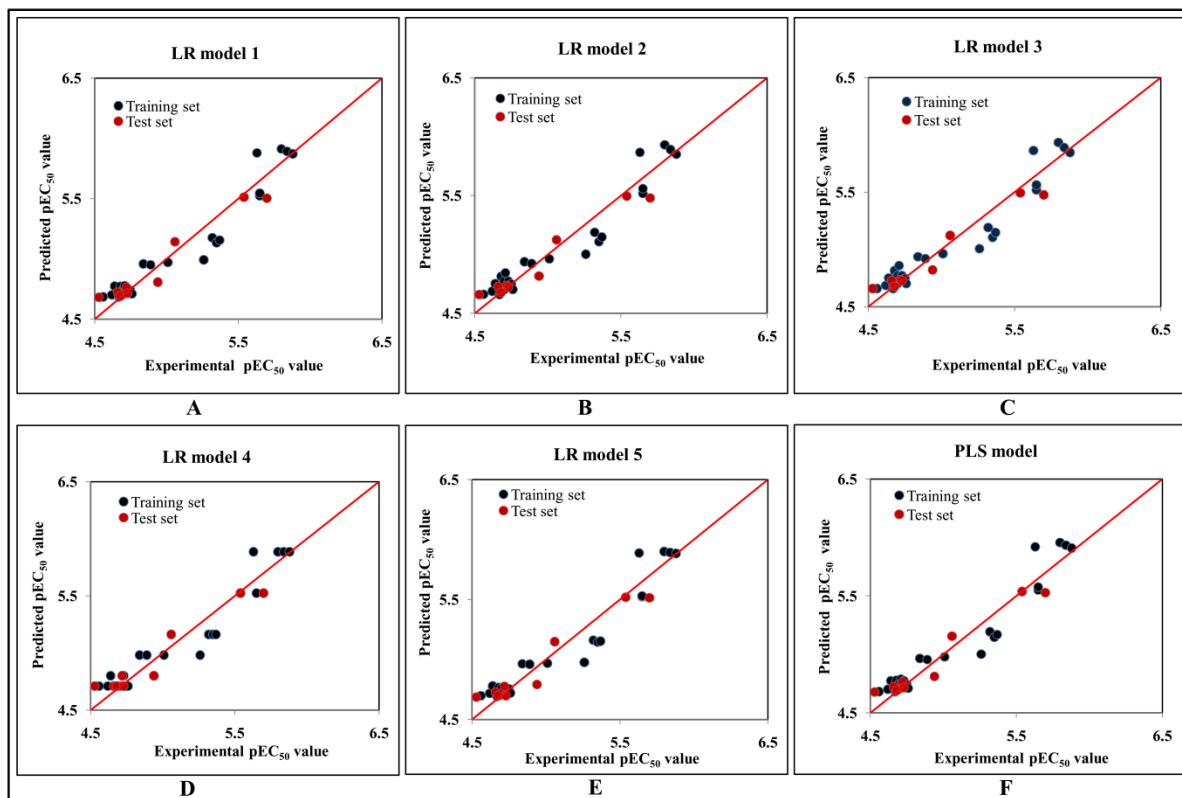


Figure 9. The scatter plot of the observed and the predicted cytotoxicity ($-\log EC_{50}$) values of the LR and PLS models. Figure (A-E) represents LR models and Figure F represents PLS model.

Model 1

The mono parametric equation is as follows:

$$-\log EC_{50} = 4.673(\pm 0.034) + 0.458(\pm 0.003)Eq$$

$$n_{\text{training}} = 25, R^2 = 0.926, R_{\text{adj}}^2 = 0.922, S = 0.127, F = 286.07,$$

$$\text{PRESS} = 0.440, Q^2 = 0.911, \overline{r_{\text{m(L00)}}^2} = 0.880, \Delta r_{\text{m(L00)}}^2 = 0.045, \text{MAE based criteria} \\ = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{F1}^2 = 0.937, Q_{F2}^2 = 0.931, \overline{r_{m(\text{test})}^2} = 0.828, \Delta r_{m(\text{test})}^2 = 0.049, \text{MAE based criteria} = \text{Good} \quad (4.2)$$

4.2.1. Electrochemical Equivalent (E_q)

The descriptor Electrochemical Equivalent (g/amp-hr) (E_q) of a chemical element indicates the mass of that element (in grams) transported by 1 coulomb of electric charge. Electrochemical Equivalent can be calculated as follows,

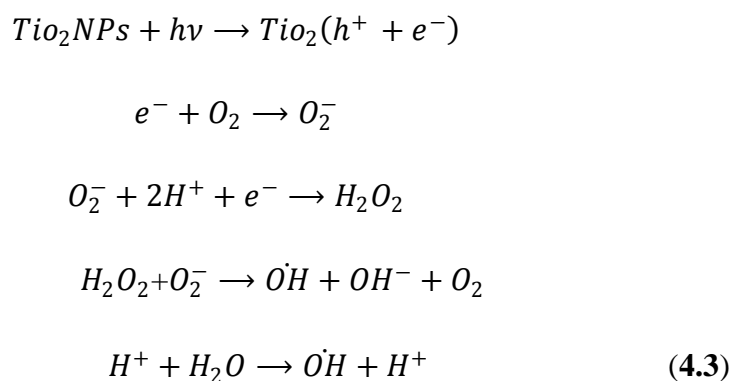
Electrochemical Equivalent = Gram molecular mass of the substance / number of electrons (involved in reaction)

This indicates that the atoms containing lower number of valence shell electrons will have higher descriptor values, as it is inversely proportional to the number of electrons in the principle valence shell taking part in the chemical reaction. The descriptor also determines the kinetics of corrosion rate and estimates the oxidizing power of metal in specific environment. The oxidizing power can be determined through the oxidation potential of a metal, which gives the measure of the likelihood of a metal to move from lower oxidation state to higher oxidation state. The transition metals can exist in different oxidation state as they have partially filled -d and -f orbital shells. The elements with less number of valence electrons will have less oxidation state and metals with less oxidation state are more harmful than the elements with higher oxidation or stable oxidation state [127]. The positive regression coefficient of this descriptor suggests that the toxicity towards hamster ovary cell will increase with an increase in the numerical value of this descriptor as shown in case of nanoparticles **6.5Ag_0.5Pt** and **6.5Ag_0.25Pt** (the Electrochemical Equivalent values are 27.06975 and 26.614825 (g/amp-hr) and their corresponding toxicity values are 5.8 and 5.84 respectively) and vice versa as shown in nanoparticles **0.05Au_0.05Pt** and **0.1Au** (the Electrochemical Equivalent values are 0.213465, 0.24496 (g/amp-hr) and their respective toxicity values are 4.67 and 4.56).

Mechanism of toxicity

The toxicity of a metal ion depends on its electrochemical features, solubility and stability. Chelating ability of the metal ion with the particular ligands of biological macro molecules

also affect the toxicity to the biological cells. Toxicity of the metal depends both quantitatively and qualitatively on the oxidation state of the metals. The lower oxidation state metals are more toxic than their higher oxidation state due to its tendency to get oxidized to form stable oxides, i.e., higher valence state hence disrupting cellular processes [128] Electron detachment from metal NPs initiates the lipid peroxidation by reactive oxygen species (ROS) such as superoxide ($O_2^{\bullet-}$) and hydroxyl radicals ($\bullet OH$) [129-130]. Using TiO_2 NPs as example, ROS is produced as per the following scheme (Eq. 4.3) in presence of light radiation:



The cellular damage as per true toxicity mechanism may involve release of metal ions. The extent of ROS production increases by direct contact of the nanoparticles (NPs) with the cell is an essential feature [131]. NPs can increase the oxidative stress as per the given mechanism by generating reactive oxygen species (ROS) which reduces the antioxidants [132] that eventually leads to cell injury and death of the cell. The production of the high energy species may attack lipids, proteins, nucleic acid or other biological macromolecules thus causing damage to the cells. They may hamper the mitochondrial structure and depolarize the membrane, even may cause impairment of the electron transport chain and activation of the NADPH (Nicotinamide adenine dinucleotide phosphate) system. Damage to the DNA may lead to cell cycle arrest and apoptosis.

Model 2

$$-\log EC_{50} = 4.643(\pm 0.036) + 0.0001(\pm 0.00001)2\chi_{pi}$$

$$n_{\text{training}} = 25, R^2 = 0.923, R_{\text{adj}}^2 = 0.920, S = 0.129, F = 277.15,$$

$$\text{PRESS} = 0.452, Q^2 = 0.909, \overline{r_{m(\text{LOO})}^2} = 0.877, \Delta r_{m(\text{LOO})}^2 = 0.047, \text{MAE based criteria} \\ = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{F1}^2 = 0.938, Q_{F2}^2 = 0.932, \overline{r_{m(\text{test})}^2} = 0.843, \Delta r_{m(\text{test})}^2 \\ = 0.042, \text{MAE based criteria} = \text{Good}$$

(4.4)

4.2.2. 2nd ionization potential ($^2\chi_{pi}$)

The next significant descriptor, 2nd ionization potential ($2\chi_{pi}$), also contributes to the cytotoxicity of the hamster ovary cell. This descriptor defines the energy needed to remove a second electron from each ion in one mole of gaseous 1^+ ion to give gaseous 2^+ ions as represented in **Eq. 4.5**.



Here, M is the atom, M^+ is the ion and e^- is the electron, i.e., ionization energy.

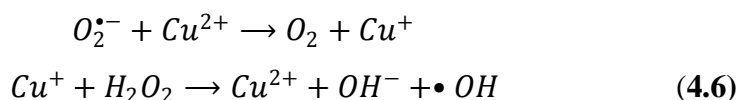
Ionization potential is the difference of energy between the ground state and state of ionization, and this amount of energy is required to completely remove the loosely attached electrons. The 2nd ionization potential is greater than 1st ionization potential and depends upon the size, charge and the type of electrons removed from outer shell of the atom. Ionization potential also determines the electronegativity and electron affinity of an atom. The less ionization energy of an atom (the energy required to remove the outer shell electron) indicated that the atom can easily lose its outer shell electron and has fewer tendencies to gain electrons. Thus, it clearly indicates that the atoms with high ionization potential will have high electronegativity. The electronegativity is responsible for the catalytic property of

the cationic form of the metal and therefore increases the cytotoxicity. The positive regression coefficient of this descriptor indicated that an atom with higher 2nd ionization potential increases the cytotoxicity of the hamster ovary cell and vice versa. As for example, the nanoparticles **6.5Ag_0.5Pt** and **6.5Ag** are highly toxic (toxicity values are 5.8 and 5.88 respectively) towards the cytotoxicity to hamster ovary cell due to their higher range of 2nd ionization potential (14350.5 and 13455 respectively), whereas in case of nanoparticles **0.25Pt** and **0.1Au**, the cytotoxicity (4.56 and 4.67 respectively) decreases with its 2nd ionization potential (447.75 and 198 kJ/mol respectively).

Mechanism of toxicity

Electronegativity depends on the atomic radius and on the formal charge of the cationic metal. Metal nanoparticles containing higher electronegativity have a tendency to gain electrons from the bonding pair of the electrons. Therefore, an increase in electronegativity suggests an increase in the catalytic properties of the cationic metal, and thus it increases the toxicity of the metal nanoparticles as described by the Haber–Weiss–Fenton cycle [133]. Electronegativity reduces with the number of valence electrons. Electronegativity of the metal separates the metal cation from the metal oxide NPs during the toxic effect. Oxidative stress caused here due to generation of intracellular ROS levels causes oxidative damage to the cells leading to apoptosis [134]. The number of ROS as OH radicals [135-136] superoxide ions [137], hydrogen peroxide (H₂O₂) is found to be responsible for the generation of oxidative stress in the cell.

The Haber–Weiss–Fenton cycle is explained using copper metal as an example (Eq.4.6):

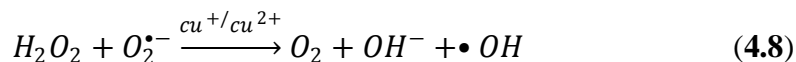


Usually $\bullet OH$ radicals are being produced in all aerobic organisms in the form of byproducts of cellular respiration as they use oxygen (molecular) to obtain energy. The problem arises when there is an imbalance between the oxidative and reductive products due to increase ROS production.



Superoxide anion radicals are products of one-electron reduction of the O₂ atom as in **Eq. 4.7**.

The electrons required for this reaction is utilized from electron transfer chain at the time of cellular respiration [138].



Thus, in the presence of the metal cation (**Eq.4.8**), the hydroxyl radical ($\bullet OH$) is formed more readily than normal. The high concentration of the hydroxyl radical becomes elevated than natural scavengers in the cell, which causes an imbalance in the antioxidants in the cell, ultimately leading to oxidative stress and cell death.

Model 3

$$-\log EC_{50} = 4.641(\pm 0.036) + 0.0013(\pm 0.0001)Rc$$

$$n_{\text{training}} = 25, R^2 = 0.923, R_{\text{adj}}^2 = 0.920, S = 0.129, F = 276.74,$$

$$\text{PRESS} = 0.452, Q^2 = 0.909, \overline{r_{m(\text{LOO})}^2} = 0.877, \Delta r_{m(\text{LOO})}^2 = 0.047, \text{MAE based criteria} \\ = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{F1}^2 = 0.938, Q_{F2}^2 = 0.932, \overline{r_{m(\text{test})}^2} = 0.842, \Delta r_{m(\text{test})}^2 = \\ 0.042, \text{MAE based criteria} = \text{Good} \quad (4.9)$$

4.2.3. Covalent radius (R_c)

The covalent radius(R_c) descriptor is a measure of the size of an atom that forms a part of one covalent bond, and it is the third common measure of the size of the atom. It is primarily calculated from the nuclear charge, i.e., atomic number and electronic configuration of the atom. The positive regression coefficient of the descriptor suggests that the numerical value of this descriptor is directly correlated with the cytotoxicity as shown in nanoparticles **6.5Ag_0.1Pt** and **6.5Ag**, where the cytotoxicity (5.63 and 5.88 respectively) increases with covalent radius of nanoparticles (956.1 (pm) and 942.5 (pm) respectively). On the other

hand, when the covalent radius of the nanoparticles decreases as in case of nanoparticles **0.25Pt** (34 pm) and **0.1Au_0.25Pt** (47.6 pm) the respective toxicity 4.67 and 4.7 also decreases. If the size of atomic radius increases, the number of shells also increases, shielding the outer electrons from the electrostatic pull of the nucleus. Again, the outer valence shells can easily lose electrons to form cation radical that may cause further modification to DNA (Deoxyribonucleic Acid) bases and enhance lipid peroxidation ultimately causes cytotoxicity.

Mechanism of toxicity

The presence of metal cations instigates the formation of sufficient amount of radicals ($\bullet\text{OH}$) than the metals do naturally. The elevated level of the reactive hydroxyl radical makes it impossible for the natural scavengers to keep the normal physiological balance in the cell. The metal cation increases the production of the free radicals both in cell and mitochondria which attack the DNA and mtDNA (mitochondrial DNA) respectively and causes fragmentation of the DNA. When metal radical attacks the protein, it causes blockade of the protein synthesis by oxidizing them leading to autocatalytic lipid peroxidation. This mechanism decreases the mitochondrial membrane potential, which leads to the loss of the mitochondrial membrane fluidity and thus the content of the matrix is spilled out into the inner membrane

Model 4

$$-\log EC_{50} = 4.709(\pm 0.338) + 0.181(\pm 0.011)\text{Amount of Ag}$$

$$n_{\text{training}} = 25, R^2 = 0.922, R_{\text{adj}}^2 = 0.919, S = 0.130, F = 272.26,$$

$$\text{PRESS} = 0.459, Q^2 = 0.907, \overline{r_{\text{m(L00)}}^2} = 0.875, \Delta r_{\text{m(L00)}}^2 = 0.048, \text{MAE based criteria} \\ = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{\text{F1}}^2 = 0.930, Q_{\text{F2}}^2 = 0.924, \overline{r_{\text{m(test)}}^2} = 0.806, \Delta r_{\text{m(test)}}^2 = \\ 0.059, \text{MAE based criteria} = \text{Good} \quad (4.10)$$

4.2.4. Amount of Ag (Ag_{amt})

The descriptor, amount of Ag, determines the measurement of silver metal concentration. The positive regression coefficient indicates that with increase in the amount of the silver metal as shown in case of nanoparticles **6.5Ag_0.5Pt** and **4.5Ag_0.5Pt** (6.5 and 4.5 mol %_respectively), the cytotoxicity (5.8 and 5.65 respectively) of the metal towards the hamster ovary cell also increases. On the other hand, when the amount of silver is reduced as shown in case of nanoparticles **0.5Ag_0.1Pt** and **1.5Ag** (0.5 and 1.5 mol % respectively) the corresponding toxicity value (4.64 and 4.89 respectively) also decreases, which clearly depicts that silver metal has a dominant role for cytotoxicity towards the hamster ovary cell. Silver metal has an antimicrobial effect and induces toxicity in many types of species [139] and chronic exposure of silver metal may cause argyria or argyrosis in humans as suggested by some authors [140]. Silver metal has a better water solubility than the other metals and thus its concentration is higher in the solution compared to other investigated metals of equal molar mass. Due to the higher concentration of silver metal available in solution, they are more toxic than the other metals.

Mechanism of toxicity

Different hypotheses have been formulated for the mechanism of silver ion toxicity to the cell. Among the various hypotheses, the silver ion release from the metal oxide [141] and generation of the reactive oxygen species (ROS) [142-143] is suggested to be most likely. Silver NPs are believed to produce toxicity through so-called Trojan-horse mechanism [142]. In this mechanism, Ag is released intracellularly after being taken up by the cell and subsequently causes death of the cell. Ag NPs accumulate into the cell and produce ROS directly or may indirectly increase ROS production by reducing the antioxidants production. Thus, it decreases the viability of the cells also induces damage of DNA and chromosomes [144-145] which ultimately leads to apoptosis of the cell. Smaller Ag NPs are more toxic than the larger ones because of their high surface to volume ratio, which further facilitates the release of the Ag ions in the cell.

Model 5

$$-\log EC_{50} = 4.682(\pm 0.035) + 0.0004(\pm 0.000026)T_c$$

$$n_{\text{training}} = 25, R^2 = 0.922, R_{\text{adj}}^2 = 0.918, S = 0.130, F = 270.44,$$

$$\text{PRESS} = 0.463, Q^2 = 0.907, \overline{r_{m(\text{LOO})}^2} = 0.874, \Delta r_{m(\text{LOO})}^2 = 0.048, \text{MAE based criteria} \\ = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{F1}^2 = 0.934, Q_{F2}^2 = 0.932, \overline{r_{m(\text{test})}^2} = 0.829, \Delta r_{m(\text{test})}^2 = \\ 0.052, \text{MAE based criteria} = \text{Good} \quad (4.11)$$

4.2.5. Thermal conductivity (T_c)

Thermal conductivity (T_c) is a property of metals. It determines the rate at which heat passes through a particular material; it is expressed as the amount of heat that passes through a unit area per unit time and possesses temperature gradient per degree per unit distance. According to the band theory, the atoms of metal crystals are very close to each other causing the orbitals to overlap each other, suggesting that there is presence of large number of electrons in a small piece of metal and due to their closeness they are referred to as bands. The filled bands are known as valence bands and partially filled bands with delocalized electrons are called conduction bands. Since in metals the closeness is very small, therefore it becomes easy for the electrons to move from valence band to conduction band. The ability of metals to conduct electricity depends on the proximity of the valence and conduction bands. Band theory also explains the possibility of the movement of delocalized electrons which is due to the overlapping of the molecular orbitals. The positive regression coefficient of this descriptor indicates that the cytotoxicity of the nanoparticles increases with its thermal conductivity and vice versa. It has been observed in case of nanoparticle **6.5Ag_0.25Pt** and **6.5Ag** that the cytotoxicity of these nanoparticles (5.84 and 5.88 respectively) increases as the thermal conductivity also increases (2806.4 (W/(m•K) and 2788.5 (W/(m•K) respectively), whereas the reverse occurs in the nanoparticles **0.25Pt** and **0.1Au** where the cytotoxicity (4.67 and 4.56 respectively) decreases with thermal conductivity (17.9 and 31.8 respectively). Thus, increase in the thermal conductivity means that there is a decrease in the band gap, which makes it easier for the movement of electrons to the conduction band

and hence overlapping of the band gaps causes oxidative stress and acute pulmonary inflammation compared to the material whose band gaps does not overlap [146-147].

Mechanism of toxicity

ROS production and oxidative stress occur due to the band gap of the nanoparticle energy band. The intracellular redox processes occurring in the biological media initiates electron transfer process from the valence band to conduction band. Burello and Worth et.al (2012)[148] stated that redox potential (E_0) of the naturally occurring reaction in the cell in context with the values of conduction band energy (E_c), and valence band energy (E_v) may be a main reason for the toxicity of the nanoparticles oxides (NPs). The overlapping of the E_c and E_v band causes oxidative stress which leads to the imbalance between the production of free radicals and the ability of the body to detoxify or counteract their harmful effects through neutralization by antioxidants. The toxicity arises due to detachment of the electron from the modified metal oxide NPs, i.e., reductive potential. Electron release in the cell interact with various molecules to produce a free radical ($\bullet OH$) that attacks the DNA double strand and blocks the replication or otherwise block the protein and oxidize them which impairs their function. When there is a sufficient DNA damage, then the cell undergoes apoptosis.

4.2.6. PLS model

After critical analysis of the statistical results (both internal and external validation parameters) obtained from the five LR models, we found that all five descriptors were significant in modeling toxicity in the Chinese hamster ovary cell. Therefore, we performed PLS regression with the same data set division using the five descriptors obtained from the LR models. The final PLS equation was developed using one latent variable.

$$-\log EC_{50} = 4.669 + 0.00918E_q + 0.00002\chi_{pi} + 0.00026R_c + 0.03627 Ag_{amt} + 0.00009 T_c$$

$$N_{train} = 25, R^2 = 0.925, R_{adj}^2 = 0.922, S = 0.127, F = 284.75, PRESS = 0.0381, Q_{(LOO)}^2 = 0.911, LV = 1, \overline{r_{m(loo)}^2} = 0.883, \Delta r_{m(loo)}^2 = 0.048$$

$$N_{test} = 9, Q_{F1}^2 = 0.944, Q_{F2}^2 = 0.938, \overline{r_{m(test)}^2} = 0.922, \Delta r_{m(test)}^2 = 0.031, MAE \text{ based criteria} = \text{Good}. \quad (4.12)$$

Using the variable importance plot (VIP) (**Figure 10**), the significance level of the descriptors was found to be in the following order: E_q , $2\chi_{pi}$, R_c , Ag_{amt} and T_c . This model could predict 91% of the response. Both external and internal validation results showed good predictivity pattern. The Q_{F1}^2 (0.944) metric also proves high predictability of the developed model. The applicability domain (AD) of the LR models was checked; it is noteworthy to mention that the test set compounds are within the AD of the developed QSTR model based on the standardization approach.

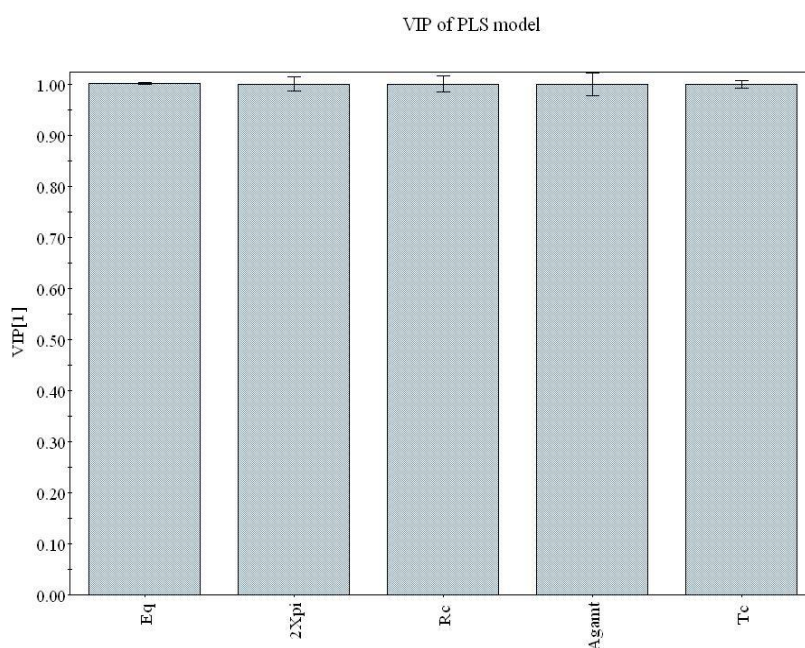


Figure 10. VIP plot of PLS model.

We have also checked the AD for the test set compounds based on the developed PLS model using the DModX approach. It was found that all the test set compounds are within the AD (D-critical=1.897) (**Figure 11**).

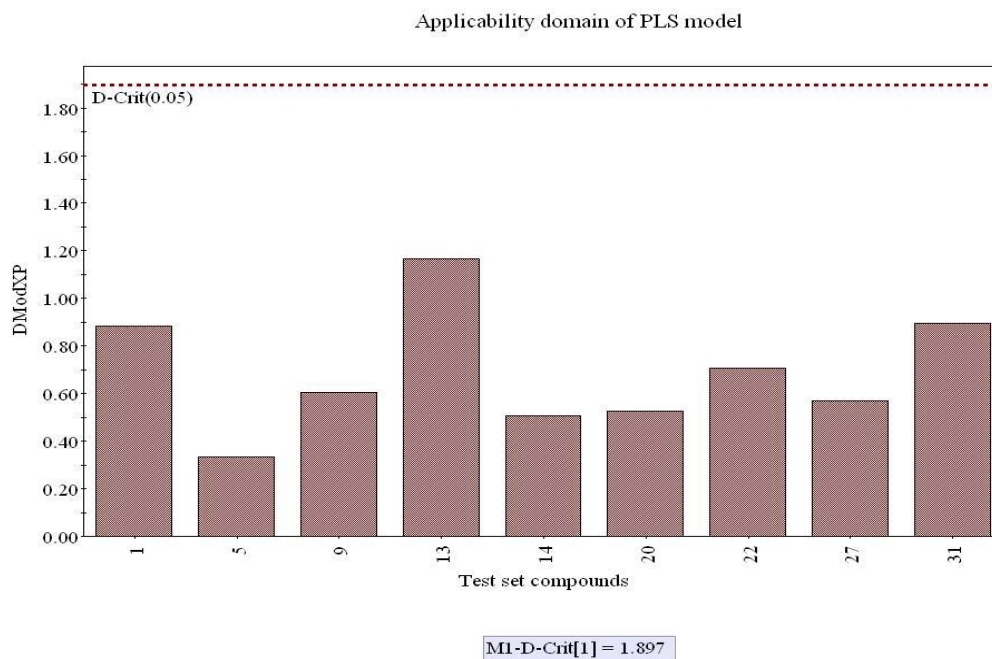


Figure 11. Applicability domain of test set compounds based on PLS model.

The various descriptor obtained in the PLS equation are elaborately explained below with the probable mode of action towards cytotoxicity to the Chinese Hamster ovary cell depicted in **Figure 12**.

The detailed mechanisms of the toxicity in terms of the descriptors are depicted in **Figure 12**.

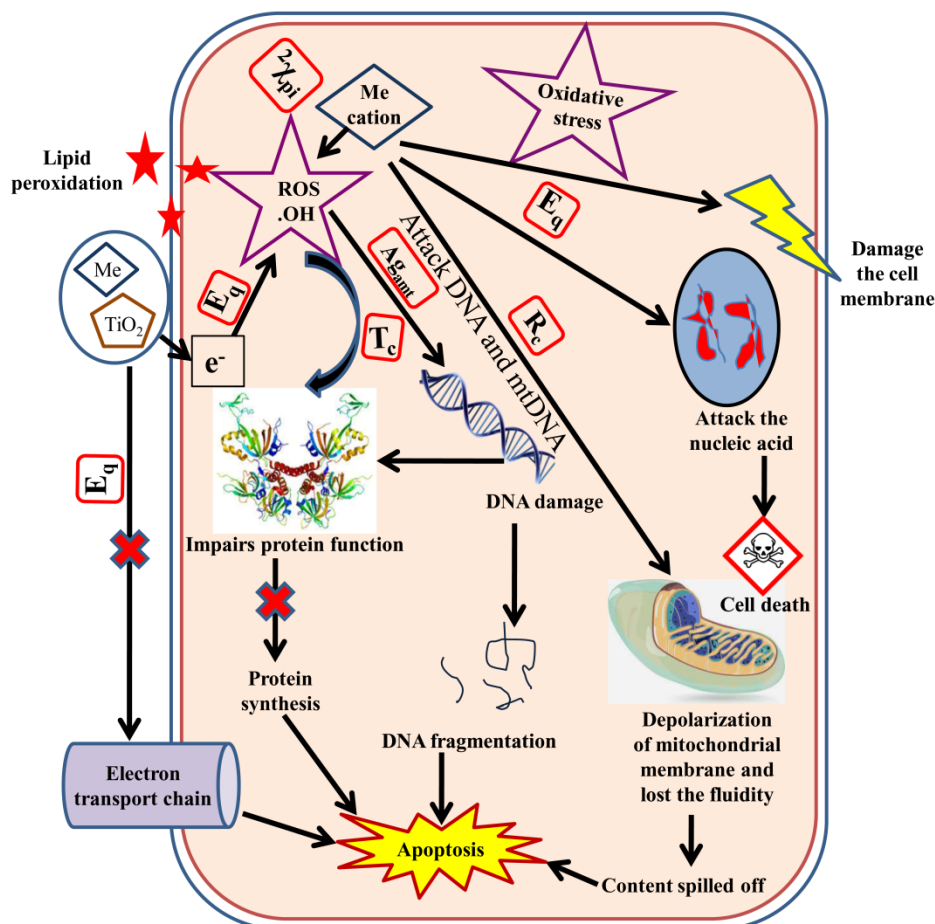


Figure 12. Mechanism of cytotoxicity of modified TiO₂ NPs towards the cell.

CHAPTER - 5

Conclusion

5. CONCLUSION

The success of any research work depends on the results obtained there from, and the conclusion drawn therein, which could bring out revealed, unrevealed or unexplored scientific explanations. These findings may further lead to develop better understanding and deep knowledge in the specific area in which the studies were performed. To meet the economically driven pressure on the chemical market in cost effective and time saving manner, computational chemistry (computer-aided drug design, molecular modeling, etc) combined with virtual screening techniques are now being hailed as the newest and fastest effective method to put new chemical entities into this chemical market. The various concepts included in the work and the different methodologies adapted are reflected in the results and explanations.

In the present study, predictive QSPR models were developed using 2D descriptors to study the various physicochemical characteristics of compounds responsible towards a certain endpoint. Determination of property of CNTs and toxicity of nanoparticles was done with the help of 2D descriptors and simple periodic table descriptors respectively. We have here implemented simple and straightforward yet robust formalism in computing descriptors, developing models, judging their prediction reliability in defined chemical space, and the diagnosed chemical information in the light of the OECD guidelines. By using QSPR models for the nanoparticles we have tried to reduce the gap between the experimental procedure and theoretical work. The models were developed using various chemometric tools and were subjected internal and external validation of the models to confirm the unbiased predictability of the developed models. The results have been fruitful and significant which leads to the conclusion that QSPR modeling can be applied for the property and toxicity determination of nanoparticles and the knowledge derived from the modeled descriptors can also be used to designing and synthesize new compounds with desired property of interest.

5.1. Predictive Quantitative Structure-Property Relationship (QSPR) Modeling for Adsorption of Organic Pollutants by Carbon Nanotubes (CNTs)

MLR strategies were employed to develop QSPR models of organic pollutants using adsorption coefficient as end point to CNTs to explore the key structural features which have

influencing effect on the adsorption of the investigatory molecules towards the MWCNTs. We have developed the models using 2D descriptors only. Prior to development of final model, different strategies for variable selection were performed to extract the most significant descriptors for the generation of the final MLR (5 models for the datasets). The extensive validation of the developed models was performed and it showed good predictability and robustness. The QSPR models were developed in compliance with the OCED principles. We have also used “Intelligent consensus predictor” tool to explore whether the quality of predictions of test set compounds can be enhanced through an “intelligent” selection of multiple MLR models. The results showed that based on the MAE based criteria, the consensus predictions of multiple MLR models are better than the results obtained from the individual models. In both the cases, the winner model is CM3. The insights obtained from the developed MLR models for dataset, we have found that (i) the descriptors like U_i , F03[O-O], F04[N-O] and ETA_BetaP influence the adsorption of organic pollutants either by π - π interactions or by strengthening π - π interactions, (ii) nArOH, F03[O-O], B03[O-O], nHBint, F04[N-O] descriptors favors the adsorption of organic pollutants through electrostatic interactions, (iii) the organic pollutants adsorbed through hydrogen bonding interactions are nArOH, F03[O-O], B03[O-O], nHBint, F04[N-O] and finally, (iv) the descriptors minsCH₃, B06[C-Cl], X0v are essential for the adsorption of organic pollutants through hydrophobic interactions. These observations were further supported by the following discussion: The organic adsorbates of CNTs were mostly aromatic compounds confirming that aromatic compounds have a better interaction with CNTs than the non aromatic pollutants due to their π electron richness and flat conformation. The systematic understanding of aromatic contaminants is therefore critical as it plays an important role in adsorption. Several studies suggested that π - π interaction is crucial for adsorption of organic compounds to CNTs [149-151] which in turn depends on size and shape of the molecules owing to the curvature of the CNTs and its substitution units. The π -system of the organic pollutants interacts with the π -system of the CNTs through π - π interaction and the interaction increases with the number of aromatic rings in the adsorbates [123] [152]. Both electron withdrawing groups (like -NO₂ and -Cl) and electron donating groups (like -NH₂, -OH) strengthen the π - π interaction between the pollutants and MWCNTs [125-126] by acting as π -electron acceptor and π -electron donor respectively. The hydroxyl group was investigated

as an electron donating substituent on adsorptive interaction among pollutants and MWCNTs, since the hydroxyls by dissociating to O^- (which has stronger electron donating ability) strengthen the $n-\pi$ electron donor acceptor (EDA) mechanism. Compounds with no aromatic ring (no π electrons) interact through hydrophobic forces. A study also suggested that CNTs act as a strong adsorbent for hydrophobic compounds due to hydrophobic interaction [153-158]. Hydroxyl groups (phenolics form) can interact through a various ways like (i) hydrophobic interaction (ii) electrostatic interaction (both attraction and repulsion) (iii) hydrogen bonding interaction and (iv) enhancing $\pi-\pi$ interaction. As the count of hydroxyls (phenolics) in the pollutants keep on increasing, the hydrophobicity decreases. Thus, it can be considered as a major factor for the adsorption of phenolics to CNTs. Hydrogen bonding can also be a major interaction between hydroxyl containing pollutants and substituted carbon nanotubes [159-160]. Hydroxyls and amino group interaction can be related to the electronic features. In one experiments it was observed that 1-naphthylamine has better adsorption on treated CNTs than the untreated one and additional observation was seen that although both 2, 4-dichlorophenol and 2-naphthol contain an -OH group, the adsorption of 2-naphthol was significant with the functionality variation of CNTs [161]. Thus, it indicates that when the adsorbates possess electronic properties, the functionality of nanotubes helps for improvement of adsorption [161]. Chen et al. (2007) [162] reported that nitro groups containing pollutants have strong adsorption than non-polar aromatic, indicating along with hydrophobic interaction there is some other essential interaction controlling adsorption comparable to π -electron polarizability related to aromatic compounds and electron donating as well as accepting property similar to compounds having more than two nitro groups. Nitro aromatic compounds besides being polar in nature, have electron accepting capacity when interacting with adsorbents having high electron polarizability property and also have high electron conjugation with π -electrons of CNTs. Thus, the higher affinity of nitro aromatic compared to other pollutants was due to $\pi-\pi$ electron donor acceptor interaction as nitrogen is a strong electron withdrawing atom, it acts as π -acceptor and carbon nanotube as π -donor [163-166]. Hydrogen bonding is also possible between nitro group of the pollutants which act as H-acceptor and functional groups substituted carbon nanotubes. Thus, we can conclude that higher number of aromatic rings, high unsaturation or electron richness of molecule, presence of polar groups substituted in aromatic ring, presence of two

oxygen atom at topological distance 3, presence of nitrogen and oxygen atoms at topological distance 4, size of the molecules and hydrophobic surface of the molecules can thus enhance the adsorption of the organic pollutants to the CNTs. On the other hand, the presence of carbon and oxygen atoms at topological distance 1 may be detrimental and can retard the adsorption of organic pollutants. Thus, this work may be helpful to remove the harmful and toxic contaminants/disposals of the by-products from the various industries making it possible to pollution free environment.

5.1. Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): A QSTR approach using simple periodic table based descriptors

Nanotechnology has a very important impact in our daily life by giving useful solution to many global problems. The influence of nanotechnology is not fully established yet. According to the reviews of recent published papers [167], there are still many gaps in the experimental data devoted to risk assessment of the nanoparticle available in today's market. The application of theoretical methods is still in its developing stage whereas the usefulness of nanoparticle is rising day by day. In this context, we have developed interpretable QSTR models and predicted the cytotoxicity of modified TiO₂ based nanoparticle towards the Chinese hamster ovary cells using simple and easily calculated periodic table descriptors and examined the applicability of such descriptor to model metal oxide nanoparticles like any other computational approaches. One of the aims of this work is to establish the simple periodic table descriptors useful for the modified nanoparticles for future use. It is believed that this type of descriptors can be used to develop the QSTR or QSAR models for other inorganic compounds also. The periodic table descriptors such as electrochemical equivalent (E_q), 2nd ionization potential ($^2\chi_{pi}$), covalent radius (R_c), amount of Ag (Ag_{amt}) and thermal conductivity (T_c) can well explain the cytotoxicity without any exhaustive calculation and thus it brings simplicity to the presented work. All the descriptors positively contributed which means that increasing the descriptor values will also increase the cytotoxicity. Oxidation number, electronegativity, molecular weight as suggested by the various periodic table based descriptors play an important role in the cytotoxicity of TiO₂ based NPs coated with various metal or mixture of metals in different concentrations. The transfer of electron from the valence band to the conduction band, the detachment of the metal cation from the

surface of the modified metal oxide surface, increase of amount of silver ion in the cell and production of radicals due to lower oxidation number induce oxidative stress, depolarization of the mitochondria impair protein function, cause fragmentation of the DNA and thus cause apoptosis and death of the cell. The major finding of the work can be summarized below:

1) *Simplicity of the proposed models:* The models are developed with the additive mixture based descriptors, which is a relatively new concept to characterize and encode the modified heterogeneous nanoparticles. The advantage of this type of descriptors is that they allow the description of the heterogeneous nanoparticles taking into account the modification on the NPs as an example of the various amount of metal on the surface of TiO₂ NPs and the variety of concentration of single metal clusters are effective in the calculation of the additive descriptors and interpretative nano-QSTR. Hence, the proposed models and approach have practical significance. The calculation of the periodic table descriptor is also not computationally demanding, and they can be easily obtained without any quantum chemical background.

2) *Mechanistic approach:* All the descriptors (E_q , ${}^2\chi_p$, R_c , Ag_{amt} and T_c) are important for the cytotoxicity of the Chinese hamster ovary cell. Metals with high electronegativity, low oxidation state, tendency to lose electrons and easy detachment of the metal cations from the modified metal oxide surface may contribute to the cell toxicity. The success of this work is to use simple descriptors for the prediction of the cytotoxicity in future of the modified metal oxide with the probable mechanistic interpretation.

3) *Cost effective and time effective:* These simple descriptors as used in this study do not involve any hard or laborious calculation thus making the use of such descriptors simple and easy. For the calculation of descriptors, there is no need to use any computational software, only the knowledge of periodic table is enough, making it both cost and time effective.

The periodic table based descriptors may thus be used to calculate the toxicity of any type of metal oxides in future studies also.

References

6. References

1. K. Thomas, P. Aguar, H. Kawasaki, J. Morris, J. Nakanishi, N. Savage, *Toxicol. Sci.* 2006, 92, 23–32
2. Roy, K., Kar, S. and Das, R.N., 2015. *A primer on QSAR/QSPR modeling: fundamental concepts*. Springer.
3. Ferreira, M.M., 2001. Polycyclic aromatic hydrocarbons: a QSPR study. *Chemosphere*, 44(2), pp.125-146.
4. Carlsen, L., Kenessov, B.N. and Batyrbekova, S.Y., 2009. A QSAR/QSTR study on the human health impact of the rocket fuel 1, 1-dimethyl hydrazine and its transformation products: Multicriteria hazard ranking based on partial order methodologies. *Environmental toxicology and pharmacology*, 27(3), pp.415-423.
5. Taylor, P.J., 1990. In Hansch, C., Sammes, PG and Taylor, JB (Eds.) *Comprehensive Medicinal Chemistry*, Vol. 4.
6. Tong, W., Hong, H., Xie, Q., Shi, L., Fang, H. and Perkins, R., 2005. Assessing QSAR limitations-A regulatory perspective. *Current Computer-Aided Drug Design*, 1(2), pp.195-205.
7. Martin, Y.C., Abagyan, R., Ferenczy, G.G., Gillet, V.J., Oprea, T.I., Ulander, J., Winkler, D. and Zefirov, N.S., 2016. Glossary of terms used in computational drug design, part II (IUPAC Recommendations 2015). *Pure and Applied Chemistry*, 88(3), pp.239-264.
8. Randić, M., 1997. On characterization of chemical structure. *Journal of chemical information and computer sciences*, 37(4), pp.672-687.
9. Segall, M., Champness, E., Obrezanova, O. and Leeding, C., 2009. Beyond profiling: using ADMET models to guide decisions. *Chemistry & biodiversity*, 6(11), pp.2144-2151.
10. Todeschini, R. and Consonni, V., 2008. *Handbook of molecular descriptors* (Vol. 11). John Wiley & Sons.
11. Livingstone, D.J., 2000. The characterization of chemical structures using molecular properties. A survey. *Journal of chemical information and computer sciences*, 40(2), pp.195-209.

12. Taylor, P.J., 1991. Quantitative drug design. the rational design, mechanistic study and therapeutic applications of chemical compounds. *Comprehensive medicinal chemistry*, 4, pp.241-294.
13. Roy, K. and Narayan Das, R., 2014. A review on principles, theory and practices of 2D-QSAR. *Current drug metabolism*, 15(4), pp.346-379..
14. Andrade, C.H., Pasqualoto, K.F., Ferreira, E.I. and Hopfinger, A.J., 2010. 4D-QSAR: perspectives in drug design. *Molecules*, 15(5), pp.3281-3294.
15. LillMA, *Drug Discov. Today* 2007, 12:1013.
16. Clark, M. and Cramer III, R.D., 1993. The probability of chance correlation using partial least squares (PLS). *Quantitative Structure-Activity Relationships*, 12(2), pp.137-145.
17. Albuquerque, M.G., Hopfinger, A.J., Barreiro, E.J. and de Alencastro, R.B., 1998. Four-dimensional quantitative structure– activity relationship analysis of a series of interphenylene 7-oxabicycloheptane oxazole thromboxane A2 receptor antagonists. *Journal of chemical information and computer sciences*, 38(5), pp.925-938.
18. VedaniA, DoblerM, *J. Med. Chem* 2002, 45:2139.
19. VedaniA, DoblerM, LillMA, *J. Med. Chem* 2005, 48:3700.
20. Hansch, C., Maloney, P.P., Fujita, T. and Muir, R.M., 1962. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature*, 194(4824), p.178.
21. Kubinyi, H., 2003. 2D QSAR Models: Hansch and Free–Wilson Analyses. In *Computational Medicinal Chemistry for Drug Discovery* (pp. 565-596). CRC Pres.
22. Krug, R.R., Hunter, W.G. and Grieger, R.A., 1976. Statistical interpretation of enthalpy–entropy compensation. *Nature*, 261(5561), p.566.
23. Taft, R.W., 1956. *Steric effects in organic chemistry*. by MS Newman, John Wiley & Sons, New York, p.556.
24. Kubinyi, H., 1979. Lipophilicity and biological acitivity. Drug transport and drug distribution in model systems and in biological systems. *Arzneimittel-Forschung*, 29(8), pp.1067-1080.
25. Franke, R., 1984. *Theoretical drug design methods* (Vol. 7). Elsevier Science Ltd.
26. T.W. Heritage, and D.R. Lowis, *Molecular hologram QSAR*, ACS Publications, 1999.
27. Fujita T, Ban T, *J. Med. Chem*1971, 14:148.

28. Hammett, L.P., 1935. Some relations between reaction rates and equilibrium constants. *Chemical Reviews*, 17(1), pp.125-136.
29. Seidell, A., 1912. A new bromine method for the determination of thymol, salicylates, and similar compounds. *Am Chem J*, 47, pp.508-526.
30. Abraham, D.J., 2003. *Burger's medicinal chemistry and drug discovery (Vol-5: chemotherapeutic agents)*. Wiley-Interscience.
31. Leonard, J.T. and Roy, K., 2006. On selection of training and test sets for the development of predictive QSAR models. *QSAR & Combinatorial Science*, 25(3), pp.235-251.
32. Roy, K., 2007. On some aspects of validation of predictive quantitative structure–activity relationship models. *Expert opinion on drug discovery*, 2(12), pp.1567-1577.
33. Siegel, R.W. and Fougere, G.E., 1995. Mechanical properties of nanophase metals. *Nanostructured Materials*, 6(1-4), pp.205-216.
34. Part, F., Berge, N., Baran, P., Stringfellow, A., Sun, W., Bartelt-Hunt, S., Mitrano, D., Li, L., Hennebert, P., Quicker, P. and Bolyard, S.C., 2018. A review of the fate of engineered nanomaterials in municipal solid waste streams. *Waste Management*, 75, pp.427-449.
35. Zeng, S., Baillargeat, D., Ho, H.P. and Yong, K.T., 2014. Nanomaterials enhanced surface plasmon resonance for biological and chemical sensing applications. *Chemical Society Reviews*, 43(10), pp.3426-3452.
36. Stephenson, C. and Hubler, A., 2015. Stability and conductivity of self assembled wires in a transverse electric field. *Scientific reports*, 5, p.15044.
37. Lyon, D. and Hubler, A., 2013. Gap size dependence of the dielectric strength in nano vacuum gaps. *IEEE Transactions on Dielectrics and Electrical Insulation*, 20(4), pp.1467-1471.
38. Valenti, G., Rampazzo, E., Bonacchi, S., Petrizza, L., Marcaccio, M., Montalti, M., Prodi, L. and Paolucci, F., 2016. Variable doping induces mechanism swapping in electrogenerated chemiluminescence of Ru (bpy) $32+$ core–shell silica nanoparticles. *Journal of the American Chemical Society*, 138(49), pp.15935-15942.
39. Kerativitayanan, P., Carrow, J.K. and Gaharwar, A.K., 2015. Nanomaterials for engineering stem cell responses. *Advanced healthcare materials*, 4(11), pp.1600-1627.

40. Senga, R., Komsa, H.P., Liu, Z., Hirose-Takai, K., Krasheninnikov, A.V. and Suenaga, K., 2014. Atomic structure and dynamic behaviour of truly one-dimensional ionic chains inside carbon nanotubes. *Nature materials*, 13(11), p.1050.
41. Medeiros, P.V., Marks, S., Wynn, J.M., Vasylenko, A., Ramasse, Q.M., Quigley, D., Sloan, J. and Morris, A.J., 2017. Single-Atom Scale Structural Selectivity in Te Nanowires Encapsulated Inside Ultranarrow, Single-Walled Carbon Nanotubes. *ACS nano*, 11(6), pp.6178-6185.
42. Vasylenko, A., Marks, S., Wynn, J.M., Medeiros, P.V., Ramasse, Q.M., Morris, A.J., Sloan, J. and Quigley, D., 2018. Electronic Structure Control of Sub-nanometer 1D SnTe via Nanostructuring within Single-Walled Carbon Nanotubes. *ACS nano*, 12(6), pp.6023-6031.
43. Medeiros, P.V., Marks, S., Wynn, J.M., Vasylenko, A., Ramasse, Q.M., Quigley, D., Sloan, J. and Morris, A.J., 2017. Single-Atom Scale Structural Selectivity in Te Nanowires Encapsulated Inside Ultranarrow, Single-Walled Carbon Nanotubes. *ACS nano*, 11(6), pp.6178-6185.
44. Vasylenko, A., Marks, S., Wynn, J.M., Medeiros, P.V., Ramasse, Q.M., Morris, A.J., Sloan, J. and Quigley, D., 2018. Electronic Structure Control of Sub-nanometer 1D SnTe via Nanostructuring within Single-Walled Carbon Nanotubes. *ACS nano*, 12(6), pp.6023-6031.
45. Eighth Nanoforum Report: Nanometrology" (PDF). Nanoforum. July 2006. pp. 13–14.
46. Lapshin, R.V., 2016. STM observation of a box-shaped graphene nanostructure appeared after mechanical cleavage of pyrolytic graphite. *Applied Surface Science*, 360, pp.451-460.
47. Wei, H. and Wang, E., 2013. Nanomaterials with enzyme-like characteristics (nanozymes): next-generation artificial enzymes. *Chemical Society Reviews*, 42(14), pp.6060-6093
48. Juzgado, A., Soldà, A., Ostric, A., Criado, A., Valenti, G., Rapino, S., Conti, G., Fracasso, G., Paolucci, F. and Prato, M., 2017. Highly sensitive electrochemiluminescence detection of a prostate cancer biomarker. *Journal of Materials Chemistry B*, 5(32), pp.6681-6687
49. DaNa. "Nanoparticles in paints". DaNa. Retrieved 28 August 201.
50. Anis, Mohab; AlTaher, Ghada; Sarhan, Wesam; Elsemary, Mona (2017). *Nanovate*. Springer. p. 105. ISBN 9783319448619.

51. Van den Borre, L. and Deboosere, P., 2017. Understanding a Man-Made Epidemic: The Relation between Historical Asbestos Consumption and Mesothelioma Mortality in Belgium. *Tijdschrift voor Sociale en Economische Geschiedenis*, 14(4), pp.116-138.
52. Anis, Mohab; AlTaher, Ghada; Sarhan, Wesam; Elsemary, Mona (2017). *Nanovate*. Springer. p. 105. ISBN 9783319448619.
53. Kašpar, J., Fornasiero, P. and Graziani, M., 1999. Use of CeO₂-based oxides in the three-way catalysis. *Catalysis Today*, 50(2), pp.285-298.
54. World Health Organization, 2017. WHO guidelines on protecting workers from potential risks of manufactured nanomaterials. World Health Organization.
55. Truskewycz, A., Patil, S., Ball, A. and Shukla, R., 2018. Iron Nanoparticles for Contaminated Site Remediation and Environmental Preservation. In *Nanobiotechnology* (pp. 323-373). CRC Press.
56. Apul, O.G., Wang, Q., Shao, T., Rieck, J.R. and Karanfil, T., 2012. Predictive model development for adsorption of aromatic contaminants by multi-walled carbon nanotubes. *Environmental science & technology*, 47(5), pp.2295-2303.
57. Apul, O.G., Xuan, P., Luo, F. and Karanfil, T., 2013. Development of a 3D QSPR model for adsorption of aromatic compounds by carbon nanotubes: comparison of multiple linear regression, artificial neural network and support vector machine. *RSC Advances*, 3(46), pp.23924-23934.
58. Kar, S., Gajewicz, A., Roy, K., Leszczynski, J. and Puzyn, T., 2016. Extrapolating between toxicity endpoints of metal oxide nanoparticles: Predicting toxicity to *Escherichia coli* and human keratinocyte cell line (HaCaT) with Nano-QTTR. *Ecotoxicology and environmental safety*, 126, pp.238-244.
59. De, P., Kar, S., Roy, K. and Leszczynski, J., 2018. Second generation periodic table-based descriptors to encode toxicity of metal oxide nanoparticles to multiple species: QSTR modeling for exploration of toxicity mechanisms. *Environmental Science: Nano*, 5(11), pp.2742-2760.
60. Puzyn, T., Rasulev, B., Gajewicz, A., Hu, X., Dasari, T.P., Michalkova, A., Hwang, H.M., Toropov, A., Leszczynska, D. and Leszczynski, J., 2011. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles. *Nature nanotechnology*, 6(3), p.175.

61. Gajewicz, A., Schaeublin, N., Rasulev, B., Hussain, S., Leszczynska, D., Puzyn, T. and Leszczynski, J., 2015. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies. *Nanotoxicology*, 9(3), pp.313-325.
62. Islam, N. and Miyazaki, K., 2010. An empirical analysis of nanotechnology research domains. *Technovation*, 30(4), pp.229-237.
63. Wonders, S., 2002. *Endless Frontiers: A Review of the National Nanotechnology Initiative*. National Research Council, Washington DC, 5.
64. Vaseashta, A., Vaclavikova, M., Vaseashta, S., Gallios, G., Roy, P. and Pummakarnchana, O., 2007. Nanostructures in environmental pollution detection, monitoring, and remediation. *Science and Technology of Advanced Materials*, 8(1-2), p.47.
65. Hansen, S.F., Maynard, A., Baun, A. and Tickner, J.A., 2008. Late lessons from early warnings for nanotechnology. *Nature nanotechnology*, 3(8), p.444.
66. Cattaneo, A.G., Gornati, R., Sabbioni, E., Chiriva-Internati, M., Cobos, E., Jenkins, M.R. and Bernardini, G., 2010. Nanotechnology and human health: risks and benefits. *Journal of applied Toxicology*, 30(8), pp.730-744.
67. Scida, K., Stege, P.W., Haby, G., Messina, G.A. and García, C.D., 2011. Recent applications of carbon-based nanomaterials in analytical chemistry: critical review. *Analytica Chimica Acta*, 691(1-2), pp.6-17.
68. Rengaraj, S., Yeon, J.W., Kim, Y. and Kim, W.H., 2007. Application of Mg-mesoporous alumina prepared by using magnesium stearate as a template for the removal of nickel: kinetics, isotherm, and error analysis. *Industrial & engineering chemistry research*, 46(9), pp.2834-2842.
69. Pan, B. and Xing, B., 2008. Adsorption mechanisms of organic chemicals on carbon nanotubes. *Environmental science & technology*, 42(24), pp.9005-9013.
70. Park, H.S. and Jun, C.H., 2009. A simple and fast algorithm for K-medoids clustering. *Expert systems with applications*, 36(2), pp.3336-3341.
71. Pope, P.T. and Webster, J.T., 1972. The use of an F-statistic in stepwise regression procedures. *Technometrics*, 14(2), pp.327-340.
72. MINITAB is a Statistical Software of Minitab Inc., USA, <http://www.minitab.com>. Pp.

73. Roy, K., Ambure, P., Kar, S. and Ojha, P.K., 2018. Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models?. *Journal of Chemometrics*, 32(4), p.e2992.
74. Artiles, M.S., Rout, C.S. and Fisher, T.S., 2011. Graphene-based hybrid materials and devices for biosensing. *Advanced drug delivery reviews*, 63(14-15), pp.1352-1360.
75. Puzyn, T., Leszczynski, J. and Cronin, M.T. eds., 2010. *Recent advances in QSAR studies: methods and applications (Vol. 8)*. Springer Science & Business Media.
76. Huang, Y.W., Cambre, M. and Lee, H.J., 2017. The toxicity of nanoparticles depends on multiple molecular and physicochemical mechanisms. *International journal of molecular sciences*, 18(12), p.2702.
77. Dreher, K.L., 2004. Health and environmental impact of nanotechnology: toxicological assessment of manufactured nanoparticles. *Toxicological Sciences*, 77(1), pp.3-5.
78. Hansen, S.F., Maynard, A., Baun, A. and Tickner, J.A., 2008. Late lessons from early warnings for nanotechnology. *Nature nanotechnology*, 3(8), p.444.
79. Cattaneo, A.G., Gornati, R., Sabbioni, E., Chiriva-Internati, M., Cobos, E., Jenkins, M.R. and Bernardini, G., 2010. Nanotechnology and human health: risks and benefits. *Journal of applied Toxicology*, 30(8), pp.730-744.
80. Mikolajczyk, A., Gajewicz, A., Mulkiewicz, E., Rasulev, B., Marchelek, M., Diak, M., Hirano, S., Zaleska-Medynska, A. and Puzyn, T., 2018. Nano-QSAR modeling for ecosafe design of heterogeneous TiO₂-based nano-photocatalysts. *Environmental Science: Nano*, 5(5), pp.1150-1160.
81. Dragon version 7, Kode srl, Milan, Italy, 2016; software available at <http://www.taletе.mi.it/index.htm>. Pp.
82. Yap, C.W., 2011. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of computational chemistry*, 32(7), pp.1466-1474.
83. *k-means clustering algorithm*; <https://sites.google.com/site/dataclusteringalgorithms/k-meansclustering-algorithm> (Accessed on 10.4.2017). Pp
84. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R. and Wu, A.Y., 2002. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (7), pp.881-892.

85. Ortega, J.P., Del, M., Rojas, R.B. and Somodevilla, M.J., 2009, March. Research issues on k-means algorithm: An experimental trial using matlab. In CEUR Workshop Proceedings: Semantic Web and New Technologies (pp. 83-96).
86. Darlington, R.B., 1990. Regression and linear models (pp. 292-293). New York: McGraw-Hill.
87. Argyle, J., 2000. Draper NR, Smith H, Applied regression analysis. STATISTICAL METHODS IN MEDICAL RESEARCH, 9, pp.517-517.
88. Bhattacharjee, A.K., Hartell, M.G., Nichols, D.A., Hicks, R.P., Stanton, B., Van Hamont, J.E. and Milhous, W.K., 2004. Structure-activity relationship study of antimalarial indolo [2, 1-b] quinazoline-6, 12-diones (tryptanthrins). Three dimensional pharmacophore modeling and identification of new antimalarial candidates. European journal of medicinal chemistry, 39(1), pp.59-67.
89. *DTC Lab web tools: http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab. Pp*
90. Snedecor, G.W., 1967. Cochran. WG, Statistical Methods. Iowa State University Press, Ames, Iowa, 196, pp.44-51.
91. Wold, S., Sjöström, M. and Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics and intelligent laboratory systems, 58(2), pp.109-130.
92. Fan, Y., 2001. shi LM, Kohn KW, pommier Y, Weinstein JN. Quantitative structure-activity relationships of camptothecin analogues: cluster analysis and genetic algorithm based studies. J Med Chem, 44, pp.3254-63.
93. Wold, S. and Sjöström, M., 1998. Chemometrics, present and future success. Chemometrics and Intelligent Laboratory Systems, 44(1-2), pp.3-14.
94. Snedecor, G.W. and Cochran, W.G., 1937. Statistical Methods-80-7*.
95. Wold, S., Eriksson, L. and Clementi, S., 1995. Statistical validation of QSAR results. Chemometric methods in molecular design, pp.309-338.
96. Pratim Roy, P., Paul, S., Mitra, I. and Roy, K., 2009. On two novel parameters for validation of predictive QSAR models. Molecules, 14(5), pp.1660-1701.
97. Consonni, V., Ballabio, D. and Todeschini, R., 2010. Evaluation of model predictive ability by external validation techniques. Journal of chemometrics, 24(3-4), pp.194-201.
98. Tropsha, A., 2010. Best practices for QSAR model development, validation, and exploitation. Molecular informatics, 29(6-7), pp.476-488.

99. Besalú, E., de Julián-Ortiz, J.V. and Pogliani, L., 2007. Trends and plot methods in MLR studies. *Journal of chemical information and modeling*, 47(3), pp.751-760.
100. Roy, K. and Mitra, I., 2012. On the use of the metric rm_2 as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. *Mini reviews in medicinal chemistry*, 12(6), pp.491-504.
101. Ojha, P.K., Mitra, I., Das, R.N. and Roy, K., 2011. Further exploring rm_2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems*, 107(1), pp.194-205.
102. Roy, K., Chakraborty, P., Mitra, I., Ojha, P.K., Kar, S. and Das, R.N., 2013. Some case studies on application of “ rm_2 ” metrics for judging quality of quantitative structure–activity relationship predictions: emphasis on scaling of response data. *Journal of computational chemistry*, 34(12), pp.1071-1082.
103. Roy, K. and Kabir, H., 2012. QSPR with extended topochemical atom (ETA) indices, 3: modeling of critical micelle concentration of cationic surfactants. *Chemical engineering science*, 81, pp.169-178.
104. Roy, K., Das, R.N., Ambure, P. and Aher, R.B., 2016. Be aware of error measures. Further studies on validation of predictive QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 152, pp.18-33.
105. Hawkins, D.M., 2004. The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), pp.1-12.
106. Schüürmann, G., Ebert, R.U., Chen, J., Wang, B. and Kühne, R., 2008. External validation and prediction employing the predictive squared correlation coefficient - Test set activity mean vs training set activity mean. *Journal of Chemical Information and Modeling*, 48(11), pp.2140-2145.
107. Consonni, V., Ballabio, D. and Todeschini, R., 2009. Comments on the definition of the Q_2 parameter for QSAR validation. *Journal of chemical information and modeling*, 49(7), pp.1669-1678.
108. Mitra, I., Saha, A. and Roy, K., 2009. Quantitative structure–activity relationship modeling of antioxidant activities of hydroxybenzalacetones using quantum chemical, physicochemical and spatial descriptors. *Chemical biology & drug design*, 73(5), pp.526-536.

109. Roy, K. and Paul, S., 2009. Exploring 2D and 3D QSARs of 2, 4-diphenyl-1, 3-oxazolines for ovicidal activity against *Tetranychus urticae*. *QSAR & Combinatorial Science*, 28(4), pp.406-425.
110. Todeschini, R. and Baccini, A., 2016. Handbook of bibliometric indicators: quantitative tools for studying and evaluating research. John Wiley & Sons.
111. Gramatica, P., 2007. Principles of QSAR models validation: internal and external. *QSAR & combinatorial science*, 26(5), pp.694-701.
112. Jaworska, J., Nikolova-Jeliazkova, N. and Aldenberg, T., 2005. QSAR applicability domain estimation by projection of the training set descriptor space: a review. *ATLA-NOTTINGHAM-*, 33(5), p.445.
113. Roy, K., Kar, S. and Ambure, P., 2015. On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems*, 145, pp.22-29.
114. Weaver, S. and Gleeson, M.P., 2008. The importance of the domain of applicability in QSAR modeling. *Journal of Molecular Graphics and Modelling*, 26(8), pp.1315-1326.
115. Yun, Y.H., Wu, D.M., Li, G.Y., Zhang, Q.Y., Yang, X., Li, Q.F., Cao, D.S. and Xu, Q.S., 2017. A strategy on the definition of applicability domain of model based on population analysis. *Chemometrics and Intelligent Laboratory Systems*, 170, pp.77-83.
116. OECD Principles for the Validation of (Q)SARs, <http://www.oecd.org/dataoecd/33/37/37849783.pdf>
117. Ambure, P., Aher, R.B., Gajewicz, A., Puzyn, T. and Roy, K., 2015. "NanoBRIDGES" software: open access tools to perform QSAR and nano-QSAR modeling. *Chemometrics and Intelligent Laboratory Systems*, 147, pp.1-13.
118. Wang, Y., Yan, F., Jia, Q. and Wang, Q., 2017. Assessment for multi-endpoint values of carbon nanotubes: Quantitative nanostructure-property relationship modeling with norm indexes. *Journal of Molecular Liquids*, 248, pp.399-405.
119. Das, S., Ojha, P.K. and Roy, K., 2017. Multilayered variable selection in QSPR: a case study of modeling melting point of bromide ionic liquids. *International Journal of Quantitative Structure-Property Relationships (IJQSPR)*, 2(1), pp.106-124.

120. Kar, S., Gajewicz, A., Puzyn, T., Roy, K. and Leszczynski, J., 2014. Periodic table-based descriptors to encode cytotoxicity profile of metal oxide nanoparticles: A mechanistic QSTR approach. *Ecotoxicology and environmental safety*, 107, pp.162-169.
121. Chirico, N. and Gramatica, P., 2011. Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient. *Journal of chemical information and modeling*, 51(9), pp.2320-2335.
122. Gotovac, S., Honda, H., Hattori, Y., Takahashi, K., Kanoh, H. and Kaneko, K., 2007. Effect of nanoscale curvature of single-walled carbon nanotubes on adsorption of polycyclic aromatic hydrocarbons. *Nano Letters*, 7(3), pp.583-587.
123. Zhang, Y., Yuan, S., Zhou, W., Xu, J. and Li, Y., 2007. Spectroscopic evidence and molecular simulation investigation of the π - π interaction between pyrene molecules and carbon nanotubes. *Journal of nanoscience and nanotechnology*, 7(7), pp.2366-2375.
124. Hall, L.H. and Kier, L.B., 1995. Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *Journal of Chemical Information and Computer Sciences*, 35(6), pp.1039-1045.
125. Star, A., Han, T.R., Gabriel, J.C.P., Bradley, K. and Grüner, G., 2003. Interaction of aromatic compounds with carbon nanotubes: correlation to the Hammett parameter of the substituent and measured carbon nanotube FET response. *Nano Letters*, 3(10), pp.1421-1423.
126. Woods, L.M., Bădescu, Ș.C. and Reinecke, T.L., 2007. Adsorption of simple benzene derivatives on carbon nanotubes. *Physical Review B*, 75(15), p.155415.
127. Walker, J.D., Enache, M. and Dearden, J.C., 2007. Quantitative cationic activity relationships for predicting toxicity of metal ions from physicochemical properties and natural occurrence levels. *QSAR & Combinatorial Science*, 26(4), pp.522-527.
128. Lovrić, J., Cho, S.J., Winnik, F.M. and Maysinger, D., 2005. Unmodified cadmium telluride quantum dots induce reactive oxygen species formation leading to multiple organelle damage and cell death. *Chemistry & biology*, 12(11), pp.1227-1234.
129. Neal, A.L., 2008. What can be inferred from bacterium-nanoparticle interactions about the potential consequences of environmental exposure to nanoparticles?. *Ecotoxicology*, 17(5), p.362.

130. Kang, S., Pinault, M., Pfefferle, L.D. and Elimelech, M., 2007. Single-walled carbon nanotubes exhibit strong antimicrobial activity. *Langmuir*, 23(17), pp.8670-8673.
131. Lin, W., Huang, Y.W., Zhou, X.D. and Ma, Y., 2006. Toxicity of cerium oxide nanoparticles in human lung cancer cells. *International journal of toxicology*, 25(6), pp.451-457.
132. Xia, T., Kovochich, M., Brant, J., Hotze, M., Sempf, J., Oberley, T., Sioutas, C., Yeh, J.I., Wiesner, M.R. and Nel, A.E., 2006. Comparison of the abilities of ambient and manufactured nanoparticles to induce cellular toxicity according to an oxidative stress paradigm. *Nano letters*, 6(8), pp.1794-1807.
133. Koppenol, W.H., 2001. The Haber-Weiss cycle—70 years later. *Redox Report*, 6(4), pp.229-234.
134. Ozben, T., 2007. Oxidative stress and apoptosis: impact on cancer therapy. *Journal of pharmaceutical sciences*, 96(9), pp.2181-2196.
135. Cho, M., Chung, H., Choi, W. and Yoon, J., 2004. Linear correlation between inactivation of *E. coli* and OH radical concentration in TiO₂ photocatalytic disinfection. *Water research*, 38(4), pp.1069-1077.
136. Sawai, J., H. Kojima, H. Igarashi, A. Hashimoto, S. Shoji, T. Sawaki, A. Hakoda, E. Kawada, T. Kokugan and M. Shimizu. 2000. "Antibacterial characteristics of magnesium oxide powder". *World J Microbiol Biotechnol* 16 (2):187-194.
137. Ames, B.N., Shigenaga, M.K. and Hagen, T.M., 1993. Oxidants, antioxidants, and the degenerative diseases of aging. *Proceedings of the National Academy of Sciences*, 90(17), pp.7915-7922.
138. Bilberg, K., Døving, K.B., Beedholm, K. and Baatrup, E., 2011. Silver nanoparticles disrupt olfaction in Crucian carp (*Carassius carassius*) and Eurasian perch (*Perca fluviatilis*). *Aquatic toxicology*, 104(1-2), pp.145-152.
139. Drake, P.L. and Hazelwood, K.J., 2005. Exposure-related health effects of silver and silver compounds: a review. *The Annals of occupational hygiene*, 49(7), pp.575-585.
140. Kittler, S., Greulich, C., Diendorf, J., Koller, M. and Epple, M., 2010. Toxicity of silver nanoparticles increases during storage because of slow dissolution under release of silver ions. *Chemistry of Materials*, 22(16), pp.4548-4554.

141. Cheng, X., Zhang, W., Ji, Y., Meng, J., Guo, H., Liu, J., Wu, X. and Xu, H., 2013. Revealing silver cytotoxicity using Au nanorods/Ag shell nanostructures: disrupting cell membrane and causing apoptosis through oxidative damage. *RSC Advances*, 3(7), pp.2296-2305.
142. Foldbjerg, R., Dang, D.A. and Autrup, H., 2011. Cytotoxicity and genotoxicity of silver nanoparticles in the human lung cancer cell line, A549. *Archives of toxicology*, 85(7), pp.743-750.
143. Lubick, N., 2008. Nanosilver toxicity: ions, nanoparticle - or both?.
144. Jiang, X., Foldbjerg, R., Miclaus, T., Wang, L., Singh, R., Hayashi, Y., Sutherland, D., Chen, C., Autrup, H. and Beer, C., 2013. Multi-platform genotoxicity analysis of silver nanoparticles in the model cell line CHO-K1. *Toxicology letters*, 222(1), pp.55-63.
145. Kim, Y.J., Yang, S.I. and Ryu, J.C., 2010. Cytotoxicity and genotoxicity of nano-silver in mammalian cell lines. *Molecular & Cellular Toxicology*, 6(2), pp.119-125.
146. Zhang, L., Zhou, P.J., Yang, F. and Wang, Z.D., 2007. Computer-based QSARs for predicting mixture toxicity of benzene and its derivatives. *Chemosphere*, 67(2), pp.396-401.
147. Nel, A., Xia, T., Meng, H., Wang, X., Lin, S., Ji, Z. and Zhang, H., 2012. Nanomaterial toxicity testing in the 21st century: use of a predictive toxicological approach and high-throughput screening. *Accounts of chemical research*, 46(3), pp.607-621.
148. Burello, Enrico, and A. P. Worth. "Development and evaluation of structure–reactivity models for predicting the in vitro oxidative stress of metal oxide nanoparticles." *Towards Efficient Designing of Safe Nanomaterials: Innovative Merge of Computational Approaches and Experimental Techniques*. Cambridge, United Kingdom: Royal Society of Chemistry (2012): 257-83.
149. Su, F. and Lu, C., 2007. Adsorption kinetics, thermodynamics and desorption of natural dissolved organic matter by multiwalled carbon nanotubes. *Journal of Environmental Science and Health, Part A*, 42(11), pp.1543-1552.
150. Wang, Z., Liu, C., Liu, Z., Xiang, H., Li, Z. and Gong, Q., 2005. π – π Interaction enhancement on the ultrafast third-order optical nonlinearity of carbon nanotubes/polymer composites. *Chemical physics letters*, 407(1-3), pp.35-39.

151. Zhang, Y., Yuan, S., Zhou, W., Xu, J. and Li, Y., 2007. Spectroscopic evidence and molecular simulation investigation of the π - π interaction between pyrene molecules and carbon nanotubes. *Journal of nanoscience and nanotechnology*, 7(7), pp.2366-2375.
152. Tournus, F., Latil, S., Heggie, M.I. and Charlier, J.C., 2005. π -stacking interaction between carbon nanotubes and organic molecules. *Physical Review B*, 72(7), p.075431.
153. Fagan, S.B., Souza Filho, A.G., Lima, J.O.G., Filho, J.M., Ferreira, O.P., Mazali, I.O., Alves, O.L. and Dresselhaus, M.S., 2004. 1, 2-Dichlorobenzene interacting with carbon nanotubes. *Nano Letters*, 4(7), pp.1285-1288.
154. Hilding, J., Grulke, E.A., Sinnott, S.B., Qian, D., Andrews, R. and Jagtoyen, M., 2001. Sorption of butane on carbon multiwall nanotubes at room temperature. *Langmuir*, 17(24), pp.7540-7544.
155. Gotovac, S., Hattori, Y., Noguchi, D., Miyamoto, J.I., Kanamaru, M., Utsumi, S., Kanoh, H. and Kaneko, K., 2006. Phenanthrene adsorption from solution on single wall carbon nanotubes. *The Journal of Physical Chemistry B*, 110(33), pp.16219-16224.
156. Yang, K., Zhu, L. and Xing, B., 2006. Adsorption of polycyclic aromatic hydrocarbons by carbon nanomaterials. *Environmental science & technology*, 40(6), pp.1855-1861.
157. Yang, K., Wang, X., Zhu, L. and Xing, B., 2006. Competitive sorption of pyrene, phenanthrene, and naphthalene on multiwalled carbon nanotubes. *Environmental science & technology*, 40(18), pp.5804-5810.
158. Zhao, J., Lu, J.P., Han, J. and Yang, C.K., 2003. Noncovalent functionalization of carbon nanotubes by aromatic organic molecules. *Applied physics letters*, 82(21), pp.3746-3748.
159. Li, X., Chen, W., Zhan, Q., Dai, L., Sowards, L., Pender, M. and Naik, R.R., 2006. Direct measurements of interactions between polypeptides and carbon nanotubes. *The Journal of Physical Chemistry B*, 110(25), pp.12621-12625.
160. Li, A., Zhang, Q., Wu, H., Zhai, Z., Liu, F., Fei, Z., Long, C., Zhu, Z. and Chen, J., 2004. A new amine-modified hypercrosslinked polymeric adsorbent for removing phenolic compounds from aqueous solutions. *Adsorption Science & Technology*, 22(10), pp.807-819.
161. Chen, W., Duan, L., Wang, L. and Zhu, D., 2008. Adsorption of hydroxyl- and amino-substituted aromatics to carbon nanotubes. *Environmental science & technology*, 42(18), pp.6862-6868.

162. Chen, W., Duan, L. and Zhu, D., 2007. Adsorption of polar and nonpolar organic chemicals to carbon nanotubes. *Environmental science & technology*, 41(24), pp.8295-8300.
163. Radovic, L.R., Moreno-Castilla, C. and Rivera-Utrilla, J., 2001. Carbon materials as adsorbents in aqueous solutions. *Chemistry and physics of carbon*, pp.227-406.
164. Hunter, C.A., 1990. Sanders. JKM The nature of π - π interactions. *J. Am. Chem. Soc*, 112, pp.5525-5534.
165. Ma, J.C. and Dougherty, D.A., 1997. The cation- π interaction. *Chemical reviews*, 97(5), pp.1303-1324.
166. Hunter, C.A., Lawson, K.R., Perkins, J. and Urch, C.J., 2001. Aromatic interactions. *Journal of the Chemical Society, Perkin Transactions 2*, (5), pp.651-669.
167. Gajewicz, A., Rasulev, B., Dinadayalane, T.C., Urbaszek, P., Puzyn, T., Leszczynska, D. and Leszczynski, J., 2012. Advancing risk assessment of engineered nanomaterials: application of computational approaches. *Advanced drug delivery reviews*, 64(15), pp.1663-1693

Appendix

Reprints



Cite this: *Environ. Sci.: Nano*, 2019, 6, 224

Predictive quantitative structure–property relationship (QSPR) modeling for adsorption of organic pollutants by carbon nanotubes (CNTs)†

Joyita Roy,‡ Sulekha Ghosh,‡ Probir Kumar Ojha and Kunal Roy *

Nanotechnology has introduced a new generation of adsorbents like carbon nanotubes (CNTs), which have drawn a widespread attention due to their outstanding ability for the removal of various inorganic and organic pollutants. The goal of this study was to develop regression-based quantitative structure–property relationship (QSPR) models for organic pollutants and organic solvents using only easily computable 2D descriptors to explore the key structural features essential for adsorption to multi-walled CNTs and improve the dispersibility index of single-walled CNTs. The statistical results of the developed models showed good quality and predictivity based on both internal and external validation metrics (dataset 1: R^2 range of 0.893–0.920, $Q_{(LOO)}^2$ range of 0.863–0.895, Q_{F1}^2 range of 0.887–0.919; dataset 2: R^2 range of 0.793–0.845, $Q_{(LOO)}^2$ range of 0.743–0.798, Q_{F1}^2 range of 0.783–0.890; dataset 3: $R^2 = 0.830$, $Q_{(LOO)}^2 = 0.775$, $Q_{F1}^2 = 0.945$). We have also tried to explore whether the quality of the predictions of test set compounds can be enhanced through an “intelligent” selection of multiple models using the “Intelligent consensus predictor” tool. The consensus results suggested that the consensus predictivity of the test set compounds gave better results than those from the individual MLR models based on different criteria (dataset 1: $Q_{F1}^2 = 0.935$, $Q_{F2}^2 = 0.935$, $MAE_{(95\%)} = \text{good}$; dataset 2: $Q_{F1}^2 = 0.887$, $Q_{F2}^2 = 0.879$, $MAE_{(95\%)} = \text{good}$). The contributed descriptors obtained from different models suggested that the organic pollutants may adsorb to the CNTs through hydrogen bonding interactions, π – π interactions, hydrophobic interactions and electrostatic interaction. Based on the observations obtained from the developed models, we have inferred that the adsorption of the organic pollutants onto the CNTs can be enhanced by the following factors: a higher number of aromatic rings, high unsaturation or electron richness of molecules, the presence of polar groups substituted in the aromatic ring, the presence of oxygen and nitrogen atoms, the size of the molecules, and the hydrophobic surface of the molecules. On the other hand, the presence of C–O groups, aliphatic primary alcohols and the presence of chlorine atoms may retard the adsorption of organic pollutants. The results also suggest that the organic solvents bearing the >N- fragment, a higher degree of branching (compactness), polar solvents with low donor number and lower ionization potential may be better solvents for enhancing the dispersibility of single-walled CNTs.

Received 22nd September 2018,
Accepted 16th November 2018

DOI: 10.1039/c8en01059e

rsc.li/es-nano

Environmental significance

Nanotechnology has introduced a new generation of adsorbents such as carbon nanotubes (CNTs), which have drawn widespread attention due to their outstanding ability for the removal of various inorganic and organic pollutants. The goal of this study was to develop quantitative structure–property relationship (QSPR) models to explore the key structural features of organic pollutants, which are essential for adsorption to multi-walled CNTs. We have also developed models to investigate the characteristics that can improve the dispersibility of single-walled CNTs. This information may be helpful in the process of removal of the harmful and toxic contaminants/disposal of the by-products from various industries by increasing the adsorption of pollutants and the dispersibility of CNTs, thus making a pollution-free environment.

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India.

E-mail: kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in;

Fax: +91 33 2837 1078; Tel: +91 98315 94140

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8en01059e

‡ These authors contributed equally.

1. Introduction

A noticeable amount of organic pollutants is released into the environment *via* various routes like the burning of fossil fuels, wastes from incineration, exhausts from automobiles, agricultural processes and industrial sectors. The disposal of

the by-products from the various industries is a challenging job for environmentalists and industries. The major problem with pollutants is their effective and safe disposal without further affecting the environment adversely. The organic pollutants (phenols, cresols, alkyl benzene sulfonates, nitro chlorobenzene, chlorinated paraffins, butadiene, synthetic dyes, insecticides, fungicides and pesticides, *etc.*) accumulate in the food chain and persist in nature and cause a significant threat to the environment.¹⁻⁴ The United States Environmental Protection Agency (EPA) has set maximum contamination levels (MCLs) and maximum contamination level goals (MCLG) for each pollutant, with no ill health effects. Sometimes the MCL level goes beyond the MCLG level because of the problem in determining small quantities of contaminants and due to lack of availability of treatment technologies and analytical methods.⁵⁻¹⁴ Thus, for the protection of the environment, the use of new and advanced materials is important. In recent years, greater focus has been placed on nanostructures as adsorbents and catalysts for removing the harmful and toxic contaminants from the environment.¹⁵⁻¹⁷ Among the various nanomaterial adsorbents, carbon nanotubes (CNTs) have been thoroughly investigated because they have a large surface area to volume ratio, inertness towards chemicals, light mass density, porous structure, great physical and chemical properties, small diameter, extraordinary optical and electrical properties, high tensile strength and efficient affinity towards pollutants. The possibility of surface modification with different functional groups makes CNTs good adsorbents¹⁸⁻²⁰ and enhances their reactivity and dispersibility for environmental protection applications.

SWCNTs have some unique mechanical, electrical and thermal properties but possess poor solubility as well as poor dispersibility in aqueous and other common organic solvents.²¹ They possess high polarizability along with van der Waals interactions and hydrophobic surface, so they are able to form aggregates with each other and with other biological and chemical systems to produce mixtures of aggregates, specifically in water.^{22,23} This bundling or entangling feature of SWCNTs causes difficulties in the dispersion of CNTs in various solvents or matrices.²⁴⁻²⁶ This also prevents the exploration of the chemistry of CNTs at a molecular level and hinders their applications²⁷ as well as limits the availability of adsorption sites for the adsorption of pollutants on the CNT surface.²⁸ The morphology variation of CNTs may also result in a difference in their aggregation tendencies, which may additionally impact their adsorption capability. The major interactions are van der Waals interactions, π - π stacking, and hydrophobic interactions for dispersibility, as suggested by many researchers.²⁹

Hyung *et al.*³⁰ reported that organic contaminants can interact with carbon nanotubes in aquatic systems and increase their stability and transport and thus, the mobility of the adsorbed organic matters on CNTs can be enhanced. The popularity of CNTs has increased since Long and Yang first reported that they can efficiently remove dioxins as compared

to activated carbon.³¹ The sorption studies performed on CNTs for metal ions³² and organic contaminants, such as butane,³³ trihalomethanes,³⁴ dioxin,³¹ xylenes,³⁵ chlorophenols,³⁶ 1,2-dichlorobenzene,³⁷ resorcinol³⁸ and polycyclic aromatic hydrocarbons (PAHs),^{15,39} suggest that CNTs can remove both organic and inorganic pollutants from water and gases.

Although a large number of pollutants are reported in the literature, adsorption data is available for only around 70 000 pollutants.⁴⁰ The determination of experimental data for a large number of pollutants is time-consuming as well as laborious and costly. The surface properties of CNTs can be modified by treating them with some active chemicals so that the CNTs do not aggregate or form bundles and hence, the dispersion of CNTs can be enhanced. QSPR modeling of organic pollutants/solvents using adsorption properties/dispersibility index by CNTs can, therefore, be of great importance for researchers and practitioners. The quantitative structure-property relationship (QSPR) approach is easier than the thermodynamic model since the input parameters of QSPR can be more easily obtained as compared to the thermodynamic models.⁴¹ QSPR not only reduces the experimental work but also predicts the features based on the chemical structures. Thus, the rationalization ideas obtained from such models provide the researchers with a conceptual framework upon which a firm discussion can be based. Recently, a great deal of work has been done with QSPR and linear surface energy relationship (LSER) modeling to develop predictive models for CNTs, including the adsorption of organic chemicals (OCs) by CNTs,⁴¹⁻⁴⁷ dispersibility of CNTs in organic solvents⁴⁸⁻⁵¹ and other properties similar to CNTs. In the past, some work has been done by researchers, for example, linear LSER models were developed by Xia *et al.*⁴³ using the biological surface index (BSAI) for the prediction and characterization of the intermolecular adsorption of OCs by CNTs. Apul *et al.*⁴⁵ reported a 3D-QSPR modeling applying the same data sets for the adsorption of aromatic compounds by CNTs and compared it with MLR, ANN and SVM methods. Another QSPR model was reported by Yilmaz *et al.*⁴⁸ using additive descriptors and quantum-chemical descriptors for the determination of the dispersibility of CNTs in different organic solvents.

The objective of the present study has been to develop statistically significant QSPR models of organic pollutants with multiple-endpoints using only easily computable 2D descriptors to explore the key structural features that are essential for adsorption to MWCNTs. We have also developed a QSPR model for organic solvents to investigate the characteristics of molecules that can improve the dispersibility of SWCNTs and may overcome the drawbacks of SWCNTs. A variable selection strategy was also employed prior to the development of final models to reduce noise in the input. We have also tried to explore whether the quality of predictions of test set compounds can be enhanced through the "intelligent" selection of multiple MLR models using an "Intelligent consensus predictor" tool.

2. Methods and materials

2.1. Dataset

We have developed QSPR models separately, using three different data sets for diverse organic contaminants with multiple-endpoints of carbon nanotubes reported in the literature.^{41,44,52} The first dataset involves the defined adsorption affinity properties (k_{∞}) of 59 organic contaminants by multi-walled carbon nanotubes (MWCNTs). The second dataset involves the adsorption affinity of 69 organic contaminants related to the specific surface area (k_{SA}) of multi-walled carbon nanotubes (MWCNTs), and the third data set involves 29 organic solvents with defined dispersibility index values (C_{max}) for single-walled carbon nanotubes (SWCNTs). We have not excluded any compound of individual data sets in our modeling analysis. All the endpoint values were taken in the logarithmic scale for the modeling purposes. The first two data sets mainly involve adsorption data for synthetic organic compounds like pyrene, naphthalene, phenol, benzene, aniline, benzoate, chloroanisole, alcohol, acetophenone, isophoron, phenanthrene dicamba, atrazine, carbamazepine, pyrimidinone, acetamide, piperidine, propionitrile, acrylic acid, thioethanol, ethanolamine, cyclopentanone, acetone and ethylene glycol derivatives, while the third data set is related to different types of solvents. The dispersibility of single-walled carbon nanotubes (SWCNTs) was measured in different solvent ranges. Here, C_{max} (mg mL^{-1}) represents the maximum dispersibility of single-walled carbon nanotubes, K_{∞} and K_{SA} are both adsorption coefficients that can be obtained from isotherm data. K_{∞} is the ratio of q_e and C_e (solid and liquid phase equilibrium concentrations, respectively, at infinite dilution conditions with an average of 0.2% aqueous solubility). K_{SA} is the normalized value of K_{∞} and the specific surface area of multi-walled carbon nanotubes (MWCNTs). The data sets are given in Tables S1, S2 and S3 in the ESI† section.

2.2. Descriptor calculation

“The molecular descriptor is the final result of a logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiments”. All the dataset compounds were drawn using the Marvin Sketch software.⁵³ The descriptors were calculated using two software tools, namely, Dragon software version 6,⁵⁴ and PaDEL-descriptor⁵⁵ software. In this work, we have calculated only 2D descriptors covering constitutional, ring descriptors, connectivity index, functional group counts, atom centered fragments, atom type E-states, 2D atom pairs, molecular properties (using Dragon software version 6) and ETA indices (using PaDEL-Descriptor software).

2.3. Data set division

Division of the dataset is a very important step for QSPR. The present work deals with three datasets containing diverse organic pollutants or solvents. In each case, all the dataset

compounds were divided into a training set and a test set using the “Modified k-medoid” clustering technique. The clustering technique categorizes a set of compounds into clusters so that the compounds present in the same cluster are similar to each other. On the other hand, when two compounds belong to two different clusters, they are said to be dissimilar in nature. The indicative compounds within a cluster are called medoids. This technique tends to select k from most middle objects or compounds as the initial medoid. Three clusters were generated for the dataset containing 59 and 29 compounds, while six clusters were generated for the dataset containing 69 compounds. We have selected approximately 25% of compounds from each data set for the test set and the remaining 75% of compounds were selected for the training set. The purpose of the training set was to develop the model and the test set was used to validate the model for prediction purposes. The same strategy was applied in the case of all three datasets for training and test set division.

2.4. Variable selection and model development

After the dataset division step, we performed data pretreatment to remove intercorrelated descriptors from all three sets of datasets. Prior to the development of final models, we tried to extract the important descriptors from the large pool of initial descriptors using various variable selection strategies.^{56,57} In case of the dataset containing 59 and 69 organic pollutants, we separately ran a stepwise regression and selected some descriptors in each case. After removing the selected descriptors obtained from the first stepwise regression run, we ran the stepwise regression again using the remaining pool of descriptors, and we repeated the same procedure. In this way, we selected some manageable numbers of descriptors and made a reduced pool of descriptors. In the case of the dataset containing 29 compounds, we developed GA equations and made a descriptor pool using the descriptors obtained from the GA (genetic algorithm) equations. After that, we ran the best subset selection for all three datasets using the reduced pools of descriptors. For this, we used a tool developed in our laboratory.⁵⁸ Five (three models were selected) and four (two models were selected) descriptor models were generated in the case of the dataset containing 59 organic pollutants, whereas six (three models were selected) and five (two models were selected) descriptor models were generated for the dataset containing 69 organic pollutants. Among the equations generated from the best subset selection, we selected five models, five models and four models for 59, 69 and 29 compounds, respectively, based on MAE criteria.⁵⁹ Descriptors were selected from the GA and stepwise regression models and a descriptor pool was generated. Finally, the selected models were run using the intelligent consensus predictor (ICP) tool developed in our laboratory⁶⁰ to explore whether the quality of predictions of external compounds could be enhanced through an “intelligent” selection of multiple models (in this report, five models were selected).

The multilayered strategies like data pretreatment,⁵⁸ stepwise regression,⁶¹ genetic method⁶² and best subset

selection⁵⁸ were involved for the selection of variables prior to the development of the final models and different steps are discussed separately in the ESI† section.

2.4.1. Intelligent consensus predictor (ICP).⁶⁰ This software was used to judge the performance of consensus predictions in comparison to their quality obtained from the individual (MLR) models based on the MAE based criteria (95%). It is obvious that a single model might not be equally useful for prediction for the whole test set compounds, which means that one QSPR model may be the best model for prediction of a test compound while the other model may be the best predictor for another test compounds. For this reason, we have selected five models in the case of a dataset containing 59 (M1–M5) and 69 (N1–N5) organic contaminants, and performed consensus prediction using the “Intelligent consensus predictor” tool to explore whether the quality of the predictions of the test set compounds could be enhanced through an “intelligent” selection of multiple models. The steps involved in the development of both MLR and PLS models are represented schematically in Fig. 1.

2.5. Statistical validation metrics

In order to judge the predictivity and reliability of the developed QSPR models, we have examined the statistical quality, applying both internal and external validation metrics. In this work, we have used various statistical parameters like determination coefficient R^2 , explained variance R^2_{adj} , variance ratio (F), and standard error of estimate (s). These parameters are

not sufficient to evaluate the predictive potential of the model, so we have used some other classical parameters for validation of the models. The internal predictivity parameters like the leave-one-out cross-validated correlation coefficient (Q^2_{LOO}), and external predictivity parameters like R^2_{pred} or Q^2_{F1} , Q^2_{F2} and concordance correlation coefficient (CCC), were also calculated. We also calculated some r^2_{m} parameters like $r^2_{\text{m(LOO)}}$ and $\Delta r^2_{\text{m(LOO)}}$ for internal validation and $r^2_{\text{m(test)}}$ and $\Delta r^2_{\text{m(test)}}$ for external validation.⁶³ The basic objective of the predictive performance of QSPR models is to investigate the prediction errors of an external set, which should be within the chemical and response-based domain of the internal set (*i.e.*, training set). The Q^2_{ext} -based metrics (*i.e.*, R^2_{pred} and Q^2_{F2}) are not always able to provide the correct indication of the prediction quality because of the influence of the response range as well as the distribution of the values of response in both the training and test set compounds.⁵⁹ Thus, we have also validated the models using the mean absolute error (MAE) criteria for both external and internal validation.⁵⁹ The error based metrics were used to determine the true indication of the prediction quality in terms of prediction error since they do not evaluate the performance of the model in comparison with the mean response (Roy *et al.*, 2016 (ref. 59)). The threshold values of Q^2 , Q^2_{F2} , R^2_{pred} , $r^2_{\text{m(test)}}$, $r^2_{\text{m(LOO)}}$ are 0.5 and for CCC, it is 0.750.^{64,65} The limit for $\Delta r^2_{\text{m(test)}}$ and $\Delta r^2_{\text{m(LOO)}}$ is 0.2. Recently, Roy *et al.* reported that a single model might not be equally useful in the prediction for the whole test set compounds, *i.e.*, one QSPR model may be the best model for prediction of a test compound while the other

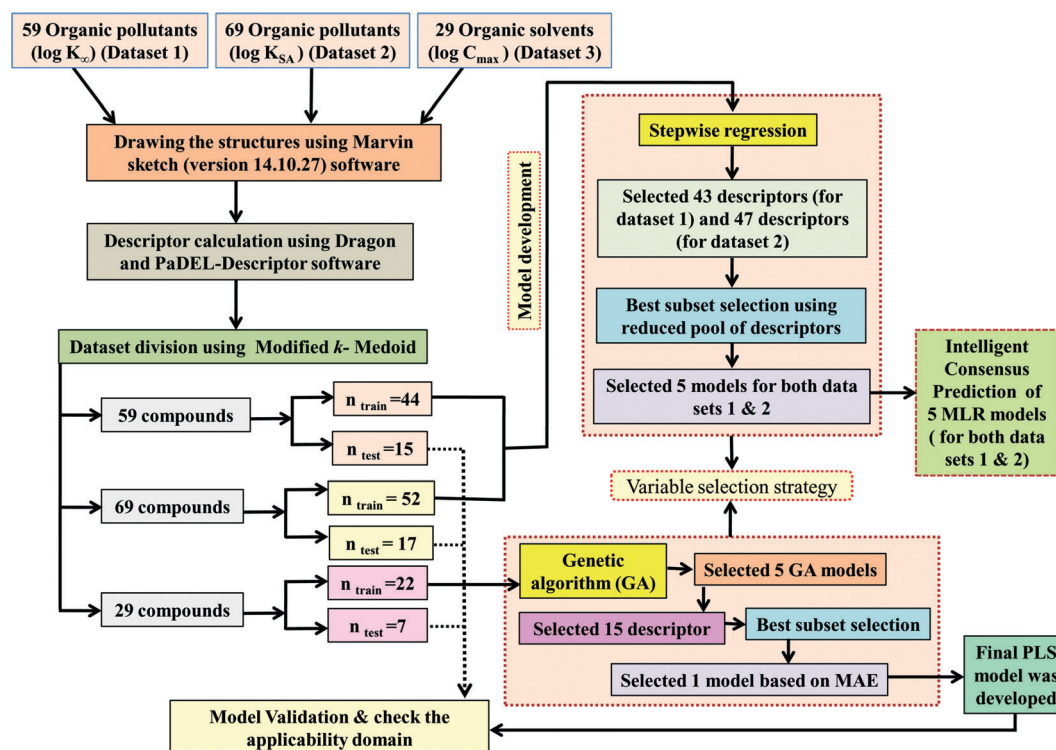


Fig. 1 Schematic representation of the steps involved in the development of QSPR models.

model may be the best predictor for another test compound. For this reason, we have also performed Intelligent consensus prediction (ICP) using multiple QSPR models to determine whether the quality of the predictions of test set compounds can be enhanced through an “intelligent” selection. Here, a simple average of predictions from all the models is not considered; only ‘qualified models’ are taken into account.

2.6. Applicability domain

“The applicability domain of a (Q)SAR is the physicochemical, structural, or biological space, knowledge or information on which the training set of the model has been developed, and for which it is applicable to make predictions for new compounds. The applicability domain of a (Q)SAR should be described in terms of the most relevant parameters, i.e., usually those that are descriptors of the model. Ideally, the (Q)SAR should only be used to make predictions within that domain by interpolation not extrapolation”. The AD of the QSAR model is characterized by the molecular properties of the training set compounds. The AD criteria help to check whether the test/query compound under consideration is inside the AD or not. Here, we have checked the applicability domain of test set compounds of the developed models, employing the standardization approach (for first two data sets) using the software developed in our laboratory⁶⁶ and a DModX (distance to model X) approach⁶⁷ at 99% confidence level using SIMCA-P software⁶⁸ (for the third data set). The predictability of a QSPR model is good if the molecules are present within the domain of the chemical space of the training set molecules.

2.7. Software used

Marvin Sketch version 5.5.0.1 (ref. 53) was used to draw chemical structures. Descriptors were calculated by the PADEL-Descriptor software⁵⁵ and Dragon software version 6.⁵⁴ Clustering of each data set was done by the “Modified K-Medoid” tool version 1.3 (ref. 58) for its splitting into a training set and a test set. Data Pretreatment version 1.2 was used to remove intercorrelated descriptors. Stepwise regression analysis was done by the MINITAB software version 13.14.⁶⁹ Genetic Algorithm was done by using the Genetic Algorithm tool version 4.1.⁵⁸ Best subset selection⁵⁸ and intelligent consensus predictor tool⁶⁰ were used to generate the QSPR models.

3. Results and discussion

We have developed QSPR models (five MLR models for each of the datasets containing 59 and 69 organic contaminants, and one PLS model for the dataset containing 29 organic contaminants) for three datasets containing diverse organic pollutants with defined adsorption affinities for MWCNTs (for datasets 1 and 2), and the dispersibility index of SWCNTs (for dataset 3), using reduced descriptors pools obtained by different strategies as discussed in the Materials and methods section. We checked the statistical quality of all the

individual models using both internal and external validation parameters, which showed that the models are statistically significant (Table 1). We also checked the MAE-based criteria for all the models.⁵⁹ All the models passed the MAE-based criteria.⁵⁹ Besides the routinely used validation parameters, we also checked the consensus predictions (for datasets 1 and 2 only) using the developed MLR models employing a newly developed “Intelligent consensus predictor” tool⁶⁰ to check whether the quality of the predictions of the test set compounds can be enhanced through an “intelligent” selection of multiple MLR models. We found that the consensus predictions of multiple MLR models are better (based on MAE based criteria) than the results obtained from the individual models as shown in Table 1 (here, in both cases, the winner model is CM3). It was also found that the consensus predictions of the test set compounds are better as compared to the individual MLR models based on not only the MAE-based criteria but also the other external validation metrics used in this work as shown in Table 1. All the individual models are mentioned below and the descriptors are discussed elaborately. In the equation, n_{training} is the number of compounds used to develop the models and n_{test} is the number of compounds used for the external prediction of the developed models. The values of leave-one-out (LOO) cross-validated correlation coefficient (Q^2) (Q^2 in the range of 0.863–0.895 for dataset 1; 0.743–0.798 for data set 2 and 0.775 for dataset 3) above the critical value of 0.5 signify the statistical reliability of the models. The predictability of the models was judged by means of predictive R^2 (R^2_{pred}) or Q^2_{F1} (Q^2_{F1} range of 0.887–0.919 for dataset 1; 0.783–0.890 for data set 2 and 0.945 for dataset 3) and Q^2_{F2} (Q^2_{F2} range of 0.886–0.919 for dataset 1; 0.768–0.882 for data set 2 and 0.938 for dataset 3), which show the good predictive ability of the models. The statistical results of all the models are summarized in Table 1. The PLS model developed from dataset 3 was also validated using a randomization test through randomly reordering (100 permutations) the dependent variable ($\log C_{\text{max}}$) using the SIMCA-P software.⁶⁸ Here, the intercept values for both R^2 and Q^2 are below the stipulated values ($R^2_{\text{int}} < 0.4$ and $Q^2_{\text{int}} < 0.05$), which confirmed that the developed model was not obtained by chance (Fig. S1 in ESI†). We have also checked the intercorrelation among the modeled descriptors for MLR models based on the Pearson correlation coefficient using the SPSS software.⁷⁰ The results showed that there is no intercorrelation between the modeled descriptors.

From the observations obtained from the modeled descriptors, it has been found that the organic pollutants may interact with the MWCNTs through different mechanisms like hydrogen bonding interactions, hydrophobic interactions, π - π interactions and electrostatic interactions as discussed below.

3.1. Dataset 1 : 59 organic pollutants

The significant descriptors obtained from the five MLR models (see Models M1–M5) for the adsorption properties

Table 1 Statistical quality and validation parameters obtained from the developed MLR and PLS models

Dataset	Type of model	Training set statistics				Test set statistics									
		R^2	Model $Q^2_{(LOO)}$	MAE_train	$r^2_{m(LOO)}$	$\Delta r^2_{m(LOO)}$	R^2_{pred} Q^2_{FI}	Q^2_F	CCC	$r^2_{m(est)}$	$\Delta r^2_{m(test)}$	MAE (100%)	MAE (95%)	MAE	
59 organic contaminants	Individual models (M1–M5)	IM1	0.920	0.895	Good	0.851	0.078	0.887	0.886	0.934	0.745	0.104	0.271	0.240	Good
		IM2	0.912	0.892	Good	0.848	0.079	0.916	0.915	0.952	0.817	0.072	0.221	0.197	Good
		IM3	0.905	0.880	Good	0.832	0.075	0.919	0.919	0.954	0.825	0.069	0.213	0.189	Good
		IM4	0.893	0.872	Good	0.821	0.092	0.918	0.917	0.953	0.806	0.074	0.213	0.187	Good
		IM5	0.893	0.863	Good	0.808	0.086	0.915	0.914	0.950	0.798	0.076	0.222	0.199	Good
Consensus models	CM0	—	—	—	—	—	0.917	0.916	0.952	0.800	0.074	0.227	0.203	Good	
	CM1	—	—	—	—	—	0.917	0.916	0.952	0.800	0.074	0.227	0.203	Good	
	CM2	—	—	—	—	—	0.919	0.919	0.953	0.803	0.073	0.221	0.196	Good	
69 organic contaminants	Individual models (N1–N5)	CM3	—	—	—	—	0.935	0.935	0.962	0.812	0.059	0.187	0.163	Good	
		IM1	0.845	0.798	Moderate	0.709	0.087	0.809	0.795	0.908	0.783	0.048	0.319	0.271	Moderate
		IM2	0.842	0.790	Moderate	0.723	0.114	0.830	0.818	0.918	0.805	0.050	0.359	0.323	Good
	IM3	0.842	0.788	Good	0.714	0.081	0.783	0.768	0.890	0.712	0.140	0.340	0.265	Good	
	IM4	0.829	0.785	Good	0.709	0.087	0.812	0.799	0.903	0.748	0.044	0.330	0.286	Moderate	
Consensus models	IM5	0.793	0.743	Good	0.709	0.087	0.890	0.882	0.940	0.836	0.090	0.273	0.247	Good	
	CM0	—	—	—	—	—	0.862	0.852	0.929	0.818	0.002	0.284	0.245	Good	
	CM1	—	—	—	—	—	0.862	0.852	0.929	0.818	0.002	0.284	0.245	Good	
29 organic contaminants	CM2	—	—	—	—	—	0.865	0.855	0.930	0.820	0.014	0.279	0.241	Good	
	CM3	—	—	—	—	—	0.887	0.879	0.941	0.851	0.040	0.263	0.235	Good	
	P1	0.830	0.775	Good	0.689	0.115	0.945	0.938	0.991	0.909	0.048	0.152	—	Good	

CM0 = Ordinary consensus predictions. CM1 = Average of predictions from individual models IM1 through IM5. CM2 = Weighted average predictions from individual models IM1 through IM5. CM3 = Best selection of predictions (compound-wise) from individual models IM1 through IM5. *Note that we have run the "Intelligent consensus predictor tool" using the options, AD: No; Dixon Q-test: No; Euclidean distance: No.

($\log K_{oc}$) of 59 organic chemicals on MWCNTs are X0v, nArOH, B01[C-O], B06[C-Cl], Ui, F03[O-O], F04[N-O], ETA_BetaP, minsCH₃, B03[O-O] and nHBint4, which regulate the adsorption properties of the organic pollutants. The contribution of the descriptors can be easily identified from the regression coefficient of the independent variables. In this case, all the descriptors contributed positively (positive regression coefficients), except the B01[C-O] descriptor (negative regression coefficient). The definition, contribution and frequency of the contributed descriptors are shown in Table S4 in the ESI.† We have checked the applicability domain of the developed MLR models using the standardization approach to confirm whether there is any compound present outside the applicability domain or not. It was found that one compound (compound number 41) for model M1 is situated outside the applicability domain, while compound number 56 is situated outside the domain of applicability in case of models M2, M3, M4 and M5; however, these compounds showed good predictivity based on the models. The scatter plot of the observed vs. predicted adsorption coefficient for all the MLR models are shown in Fig. 2.

$$\begin{aligned} \text{Model M1. } \log k_{oc} = & -4.62(\pm 0.337) + 0.834(\pm 0.155) \times \text{Ui} \\ & + 0.663(\pm 0.220) \times \text{B06[C-Cl]} \\ & + 0.641(\pm 0.057) \times \text{X0v} \\ & + 0.600(\pm 0.091) \times \text{nArOH} \\ & - 0.611(\pm 0.121) \times \text{B01[C-O]} \end{aligned}$$

$$n_{\text{training}} = 44, R^2 = 0.920, R_{\text{adj}}^2 = 0.908, S = 0.294, F = 85.93,$$

$$\text{PRESS} = 4.267, Q^2 = 0.895, \overline{r_{m(\text{L100})}^2} = 0.851,$$

$$\Delta r_{m(\text{L100})}^2 = 0.078, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 15, Q_{F1}^2 = 0.887, Q_{F2}^2 = 0.886, \overline{r_{m(\text{test})}^2} = 0.745, \Delta r_{m(\text{test})}^2 = 0.104,$$

$$\text{CCC} = 0.934, \text{MAE} = \text{Good}$$

$$\begin{aligned} \text{Model M2. } \log k_{oc} = & -8.51(\pm 0.722) + 0.803(\pm 0.048) \times \text{X0v} \\ & + 0.681(\pm 0.146) \times \text{F03[O-O]} \\ & + 0.415(\pm 0.144) \times \text{F04[N-O]} \\ & + 3.27(\pm 0.491) \times \text{ETA_BetaP} \\ & + 0.204(\pm 0.067) \times \text{minsCH}_3 \end{aligned}$$

Experimental $\log K_{oc}$ vs predicted $\log K_{oc}$ values of 59 organic pollutants

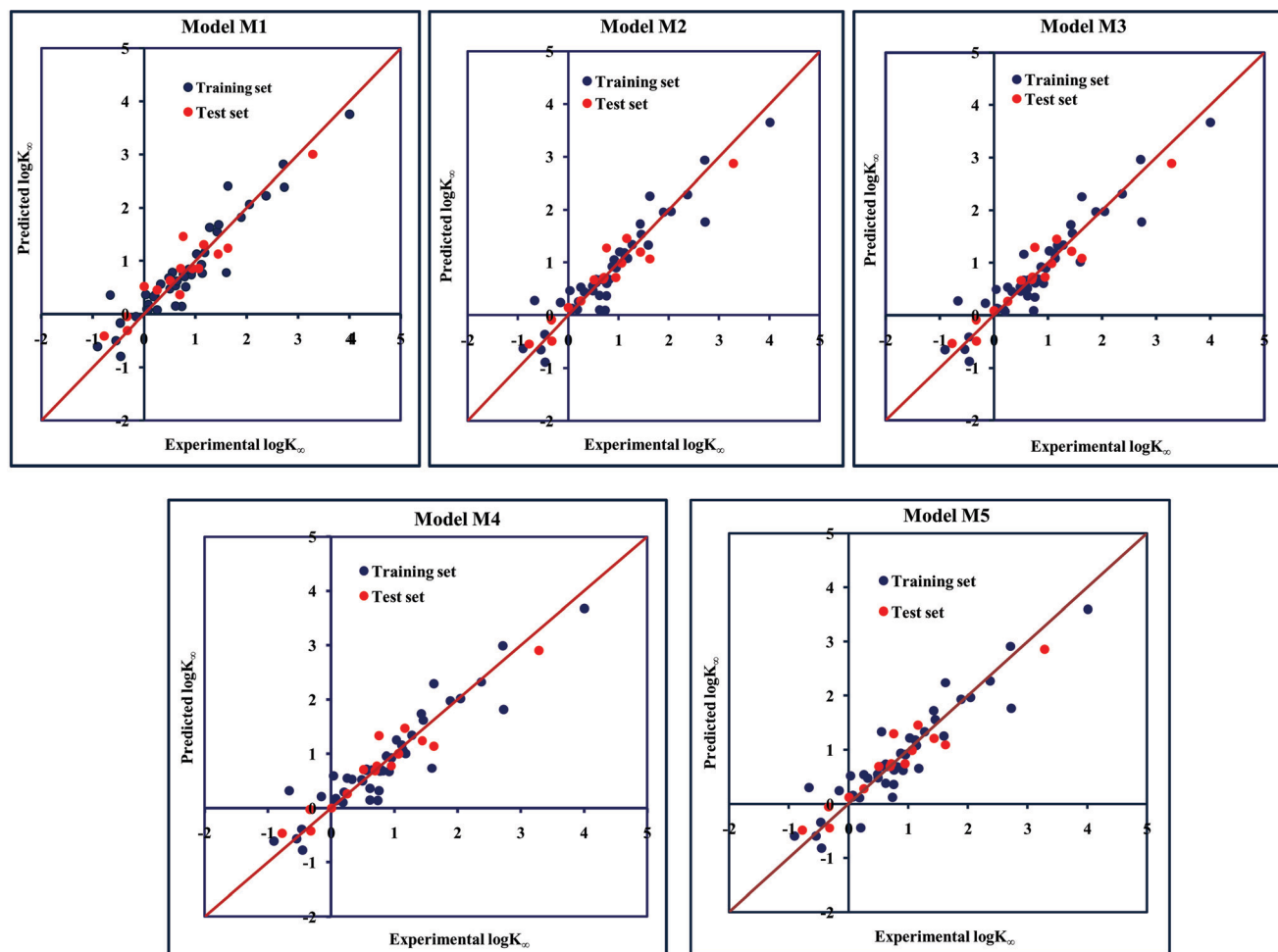


Fig. 2 The scatter plot of the observed and the predicted adsorption coefficient property ($\log K_{oc}$) of the developed MLR models (models M1–M5).

$$n_{\text{training}} = 44, R^2 = 0.912, R_{\text{adj}}^2 = 0.900, S = 0.306, F = 78.66,$$

$$\text{PRESS} = 4.356, Q^2 = 0.892, \overline{r_{\text{m(LOO)}}^2} = 0.848,$$

$$\Delta r_{\text{m(LOO)}}^2 = 0.079, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 15, Q_{\text{F1}}^2 = 0.916, Q_{\text{F2}}^2 = 0.915, \overline{r_{\text{m(test)}}^2} = 0.817, \Delta r_{\text{m(test)}}^2 = 0.072,$$

$$\text{CCC} = 0.952, \text{MAE} = \text{Good}$$

Model M3. $\log k_{\infty} = -8.68(\pm 0.746) + 0.802(\pm 0.050) \times \text{X0v}$
 $+ 0.603(\pm 0.272) \times \text{B03[O-O]}$
 $+ 3.39(\pm 0.503) \times \text{ETA_BetaP}$
 $+ 0.213(\pm 0.069) \times \text{minsCH}_3$
 $+ 0.412(\pm 0.148) \times \text{nHBint}_4$

$$n_{\text{training}} = 44, R^2 = 0.905, R_{\text{adj}}^2 = 0.893, S = 0.318, F = 72.57,$$

$$\text{PRESS} = 4.840, Q^2 = 0.880, \overline{r_{\text{m(LOO)}}^2} = 0.832,$$

$$\Delta r_{\text{m(LOO)}}^2 = 0.075, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 15, Q_{\text{F1}}^2 = 0.919, Q_{\text{F2}}^2 = 0.919, \overline{r_{\text{m(test)}}^2} = 0.825, \Delta r_{\text{m(test)}}^2 = 0.069,$$

$$\text{CCC} = 0.954, \text{MAE} = \text{Good}$$

Model M4. $\log k_{\infty} = -8.72(\pm 0.782) + 0.785(\pm 0.052) \times \text{X0v}$
 $+ 0.650(\pm 0.158) \times \text{F03[O-O]}$
 $+ 3.51(\pm 0.527) \times \text{ETA_BetaP}$
 $+ 0.202(\pm 0.073) \times \text{minsCH}_3$

$$n_{\text{training}} = 44, R^2 = 0.893, R_{\text{adj}}^2 = 0.882, S = 0.334, F = 81.11,$$

$$\text{PRESS} = 5.164, Q^2 = 0.872, \overline{r_{\text{m(LOO)}}^2} = 0.821,$$

$$\Delta r_{\text{m(LOO)}}^2 = 0.092, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 15, Q_{\text{F1}}^2 = 0.918, Q_{\text{F2}}^2 = 0.917, \overline{r_{\text{m(test)}}^2} = 0.806, \Delta r_{\text{m(test)}}^2 = 0.074,$$

$$\text{CCC} = 0.953, \text{MAE} = \text{Good}$$

Model M5. $\log k_{\infty} = -8.42(\pm 0.773) + 0.785(\pm 0.052) \times \text{X0v}$
 $+ 3.29(\pm 0.526) \times \text{ETA_BetaP}$
 $+ 0.199(\pm 0.072) \times \text{minsCH}_3$
 $+ 0.566(\pm 0.137) \times \text{nHBint}_4$

$$n_{\text{training}} = 44, R^2 = 0.893, R_{\text{adj}}^2 = 0.882, S = 0.333, F = 81.33,$$

$$\text{PRESS} = 5.543, Q^2 = 0.863, \overline{r_{\text{m(LOO)}}^2} = 0.808,$$

$$\Delta r_{\text{m(LOO)}}^2 = 0.086, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 15, Q_{\text{F1}}^2 = 0.915, Q_{\text{F2}}^2 = 0.914, \overline{r_{\text{m(test)}}^2} = 0.798, \Delta r_{\text{m(test)}}^2 = 0.076,$$

$$\text{CCC} = 0.950, \text{MAE} = \text{Good}$$

3.1.1. The descriptors related to hydrogen bonding interactions. The functional group count descriptor, $n\text{ArOH}$, repre-

sents the number of aromatic hydroxyl groups present in the compound. This descriptor influences the adsorption properties of organic pollutants by MWCNTs as indicated by its positive regression coefficient. Thus, the compounds containing a large number of aromatic hydroxyl groups may enhance the adsorption properties of organic pollutants by MWCNTs as shown in compounds 13 (pyrogallol) (containing 3-OH groups), 5 (2-phenyl phenol) (containing 1-OH group) and 14 (2,4,6 trichlorophenol) (containing 1-OH group). On the other hand, the compounds containing no aromatic hydroxyl groups are detrimental for the adsorption affinity of organic pollutants by MWCNTs as shown in compounds 18 (4-chloroaniline), 36 (benzyl alcohol) and 42 (phenethyl alcohol) (these compounds contain no aromatic hydroxyl groups). Although some compounds containing no aromatic hydroxyl groups still show high adsorption affinity for the organic pollutants by MWCNTs, it is due to some other dominating descriptors present in the model. Thus, the substitution of electron donating groups like hydroxyl groups in the aromatic ring of organic pollutants could enhance the adsorption on MWCNTs.

A 2D atom pair descriptor, F04[N-O], indicates the frequency of the N-O fragment at topological distance 4. The positive regression coefficient of the descriptor suggests that an increase in N-O fragments at topological distance 4 is directly proportional to the adsorption affinity of organic pollutants. The greater number of fragments correlates to higher adsorption properties as observed in the case of compounds 19 (2-nitroaniline) and 27 (3-nitrophenol), while the absence of such fragments at topological distance 4 has no influence on the adsorption by MWCNTs as shown in compounds 18 (4-chloroaniline), 36 (benzyl alcohol) and 42 (phenethyl-alcohol). This descriptor also indicates that the frequency of two electronegative atoms of organic pollutants (electron donating or electron withdrawing groups) should be situated at topological distance 4 for better adsorption on MWCNTs. In the case of compound number 19, nitrogen (-NH₂ group) acts as an electron donor and oxygen (-NO₂ group) acts as an electron withdrawing group, whereas in the case of compound number 27, nitrogen (-NO₂ group) acts as an electron withdrawing group, and oxygen (-OH group) acts as an electron donating group.

The E-state descriptor, nHBint4 indicates the count of potential internal hydrogen bonds separated by four edges. The positive regression coefficient suggests that hydrogen bonds of organic pollutants have the propensity to play a dominant role in enhancing the adsorption properties. Thus, the organic pollutants bearing hydrogen-bonded groups separated by four path lengths are conducive to adsorption as shown in compounds 13 (pyrogallol), 19 (2-nitroaniline) and 48 (3-chlorophenol), whereas the absence of such fragment in organic pollutants are detrimental to the adsorption affinity as shown in compounds 6 (benzene), 11 (phenol) and 42 (phenethyl alcohol).

B03[O-O] is a 2D atom pair descriptor that indicates the presence or absence of the O-O fragment at topological

distance 3. The positive regression coefficient of the descriptor indicates that the higher the frequency of this fragment, the higher is the adsorption affinity. Thus, the presence of the O–O fragment at topological distance 3 favors the adsorption of organic pollutants by MWCNTs as shown in compounds no. 12 (catechol) and 13 (pyrogallol), while compounds no. 6 (benzene), 42 (phenethyl alcohol) and 36 (benzyl alcohol) show low adsorption because these compounds have no such fragments at topological distance 3.

Hydrogen bonding is one of the key mechanisms for the adsorption of organic contaminants on CNTs. The information obtained from the descriptors $n\text{ArOH}$, $\text{F04}[\text{N-O}]$, $n\text{HBint4}$, $\text{F03}[\text{O-O}]$ and $\text{B03}[\text{O-O}]$ suggested that there may be some hydrogen bonding interactions between organic pollutants and MWCNTs, which regulate the adsorption affinity (Fig. 3) of organic pollutants toward MWCNTs. In the case of the descriptor $n\text{ArOH}$, the aromatic hydroxyl group may form hydrogen bonds with the hydroxy/carboxylic groups on the CNTs surface and the hydrogen bonds may also form between the surface-adsorbed aromatic hydroxyl group-containing organic pollutants (phenolics) and dissolved phenolics. Here, the hydroxyl group is always connected to an aromatic ring. Thus, it is obvious that this aromatic ring of organic pollutants themselves can interact with CNTs by π - π interactions. The descriptor, $\text{F04}[\text{N-O}]$, also suggested that besides the hydrogen bonding interactions, there may also be a chance to form electrostatic interactions. The electron-withdrawing groups like NO_2 may also strengthen the π - π interactions formed between the benzene derivatives (acting as π -acceptor) and CNTs (acting as π -donor). In the case of $\text{B03}[\text{O-O}]$, two oxygen atoms (hydroxyl groups) are separated by topological distance 3 and can interact with CNTs by hydrogen bonding interactions. These two electronegative atoms of organic pollutants could also interact electrostatically with CNTs and strengthen the π - π interactions formed between the organic pollutants and MWCNTs.^{39,71} It is worth noting that although the C–O bond is detrimental to the adsorption of organic pollutants on CNTs, the frequency of the O–O fragment at topological distance 3 can suppress the detrimental effect of the C–O group and influence the adsorption affinity of organic pollutants on MWCNTs. The descriptors involved in the hydrogen bonding interactions between the organic pollutants and MWCNTs are depicted in Fig. 3.

3.1.2. The descriptors related to hydrophobic interactions.

A 2D atom pair descriptor, $\text{B06}[\text{C-Cl}]$, represents the presence or absence of the C–Cl bond at topological distance 6. The positive regression coefficient of this parameter suggests that the presence of such a fragment at topological distance 6 enhances the adsorption affinity of organic pollutants towards the MWCNTs as shown in compounds 50 (4-chloroacetophenone) and 57 (2-chloronaphthene). On the other hand, compounds like 11 (phenol), 22 (4-methylphenol) and 43 (3-methylbenzyl alcohol) show poor adsorption affinity for the MWCNTs due to the absence of such a fragment.

The descriptor $X0v$ indicates a valence connectivity index of the order 0, which can be calculated through Kier and

Hall's connectivity index as shown below. This descriptor contributed positively to the adsorption affinity of organic pollutants for the MWCNTs. Thus, the size of the organic pollutants plays a crucial role in regulating the adsorption affinity of organic pollutants to MWCNTs. It has been found that on increasing the numerical value of this descriptor, the adsorption affinity of organic pollutants for MWCNTs also increases, as shown in the case of compounds 1 (pyrene), 58 (azobenzene) and 5 (2-phenyl phenol) (bigger in size), while the adsorption affinity of organic pollutants for MWCNTs decreases in the case of compounds 6 (benzene), 11 (phenol) and 36 (benzyl alcohol) (smaller in size).

The valence connectivity index of the zeroth order can be calculated by the following:

$$X0v = \sum_{i=1}^n (\delta_i^v)^{-0.5}$$

$$\delta_i^v = \frac{Z_i^v - hi}{Z_i - Z_i^v - 1}$$

In the above equation, δ_i^v = the valence vertex degree, Z_i^v = valence electrons in the i th atom, hi = the number of hydrogen atoms connected to the i th atom, Z_i = the number of electrons in the i th atom.

The E-state indices of a particular atom in a certain molecule provide information on its electronic state of that particular atom, which in turn depends on π bonds, the lone pair of electrons and δ bonds that inform the quantitative availability of the valence electrons.⁷² The descriptor minsCH_3 indicates the minimum atom type E-state CH_3 . The positive regression coefficient of this descriptor indicates that the presence of the CH_3 group has an important role in influencing the adsorption properties of organic pollutants. The numerical value of this descriptor is directly proportional to the adsorption property, which suggests that with increasing the numerical value of this descriptor, the adsorption affinity of the organic pollutants also increases as evidenced by compounds 10 (2,4-dinitrotoluene), 50 (4-chloroacetophenone) and 52 (1-methylnaphthalene). On the other hand, the adsorption affinity of organic pollutants decreases with the absence of the CH_3 group as shown in compounds 6 (benzene), 11 (phenol) and 36 (benzyl alcohol).

Hydrophobic interactions between organic pollutants and CNTs are also an important mechanism for better adsorption. The descriptors, $\text{B06}[\text{C-Cl}]$, $X0v$ and minsCH_3 suggest that the organic pollutants may be adsorbed onto the MWCNTs by hydrophobic interactions. In the case of $\text{B06}[\text{C-Cl}]$ and $X0v$, the size of the molecules (for $\text{B06}[\text{C-Cl}]$, the distance between C and Cl atoms is six, which reflects the size of the molecules) plays an important role in the adsorption affinity. The size enhances the surface area of molecules, which can regulate the hydrophobic interactions between organic pollutants and MWCNTs. The methyl group (information obtained from minsCH_3 descriptor) and CNTs are hydrophobic in nature. Thus, an increase in the minsCH_3 value



Fig. 3 Mechanistic interpretation of the descriptors related to hydrogen bonding interactions between organic pollutants and MWCNTs (dataset 1).

would indicate a higher degree of unsaturation and would enhance the reactivity. There is, therefore, a chance for hydrophobic interactions between organic pollutants and MWCNTs, which reflects better adsorption. The descriptors involved in hydrophobic interactions between organic pollutants and CNTs are depicted in Fig. 4.

3.1.3. The descriptors related to π - π interactions. The descriptor, U_i , gives information about the unsaturation index, which contributes positively to the adsorption affinity of or-

ganic pollutants by MWCNTs as indicated by the positive regression coefficient. From this descriptor, it has been suggested that the presence of unsaturated inorganic pollutants plays a crucial role in enhancing the adsorption affinity. This was demonstrated in compounds 1 (pyrene), 10 (2,4-dinitrotoluene) and 58 (azobenzene) (the numerical values of this descriptor are 3.392, 3 and 3, respectively), and *vice versa* in the case of compounds 11 (phenol), 36 (benzyl alcohol) and 42 (phenethyl alcohol) (the numerical values of this



Fig. 4 Mechanistic interpretation of the descriptors related to the hydrophobic interaction between organic pollutants and MWCNTs (dataset 1).

descriptor are 2 in each compound). Here, the compounds, 1 (pyrene), 10 (2,4-dinitrotoluene) and 58 (azobenzene) have a higher range of unsaturation index values due to the presence of a large number of double bonds.

The ETA index, ETA_BetaP, gives a measure of sigma, pi and non-bonded (*i.e.*, lone pairs capable of forming resonance with the aromatic system) electrons relative to the molecular size. Therefore, electron-richness (unsaturation) relative to the molecular size of organic pollutants is an important parameter for regulating the adsorption properties. The positive regression coefficient of this parameter indicates that the electron densities of the molecules should be higher for increasing the adsorption affinity of organic pollutants for MWCNTs, as found in compounds 1 (pyrene), 28 (1,3-dinitrobenzene) and 58 (azobenzene), whereas the compounds with low electron density show a lower range of adsorption affinities as shown in compounds 36 (benzyl alcohol), 42 (phenethyl alcohol) and 43 (3-methylbenzyl alcohol). Thus, it can be concluded that the molecules should be electron-rich for higher adsorption properties of organic pollutants.

The π - π interaction is another important mechanism involved in the adsorption of organic pollutants to CNTs. The information obtained from Ui and ETA_BetaP descriptors suggested that the organic pollutants can adsorb to MWCNTs by strong π - π interactions. The descriptors B03[O-O], F03[O-O] and F04[N-O] suggested that the [O-O] fragments at topological distance 3 and the [N-O] fragments at the topological distance 4 may strengthen the π - π interactions formed between organic pollutants and MWCNTs. The descriptor Ui suggested that unsaturation plays a crucial role for the adsorption of organic pollutants to MWCNTs. CNTs also con-

tain a large number of double bonds (unsaturation), so there is a chance to form strong π - π interactions between organic pollutants and MWCNTs, which reflects the better adsorption of these pollutants to MWCNTs; hence, a higher number of double bonds of organic pollutants enhance the adsorption affinity to MWCNTs. The descriptor, ETA_BetaP suggested that unsaturation (electron-richness) relative to the molecular size of organic pollutants plays a crucial role in regulating the adsorption properties. From this descriptor, it can be inferred that the adsorption affinity of organic pollutants to MWCNTs is increased due to the π - π interactions. The descriptors involved in π - π interactions between organic pollutants and CNTs are described graphically in Fig. 5.

3.1.4. The descriptors related to electrostatic interactions.

F03[O-O], a 2D atom pair descriptor, indicates the frequency of the O-O fragment at topological distance 3. The positive regression coefficient of this descriptor suggests that presence of a greater number of O-O bonds at the topological distance 3 might be beneficial for the adsorption affinity of organic pollutants for MWCNTs as shown in compounds 12 (catechol) and 13 (pyrogallol), whereas the opposite happens in the case of compounds 6 (benzene), 42 (phenethyl alcohol) and 43 (3-methylbenzyl alcohol) (where, no O-O fragment is present at topological distance 3). This fragment may also strengthen the π - π interactions formed between organic pollutants and MWCNTs.^{73,74} Like B03[O-O], this descriptor also suppresses the detrimental effect of the C-O group as discussed earlier in this section.

The information obtained from the descriptors, F03[O-O], B03[O-O] and F04[N-O] suggests that the organic pollutants can adhere to the surface of the MWCNTs by strong

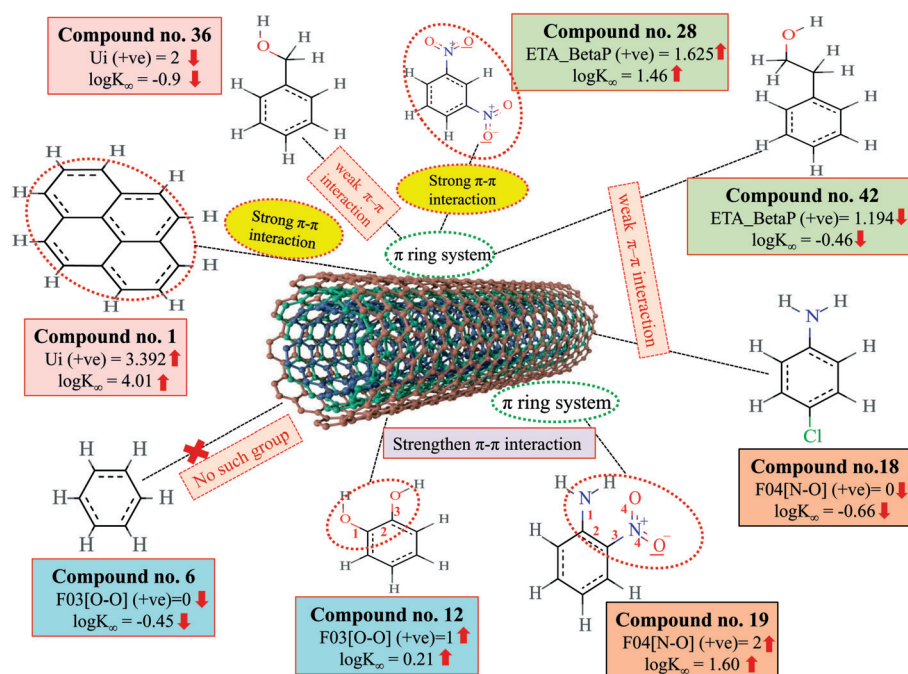


Fig. 5 Mechanistic interpretation of the descriptors related to the π - π interactions between organic pollutants and MWCNTs (dataset 1).

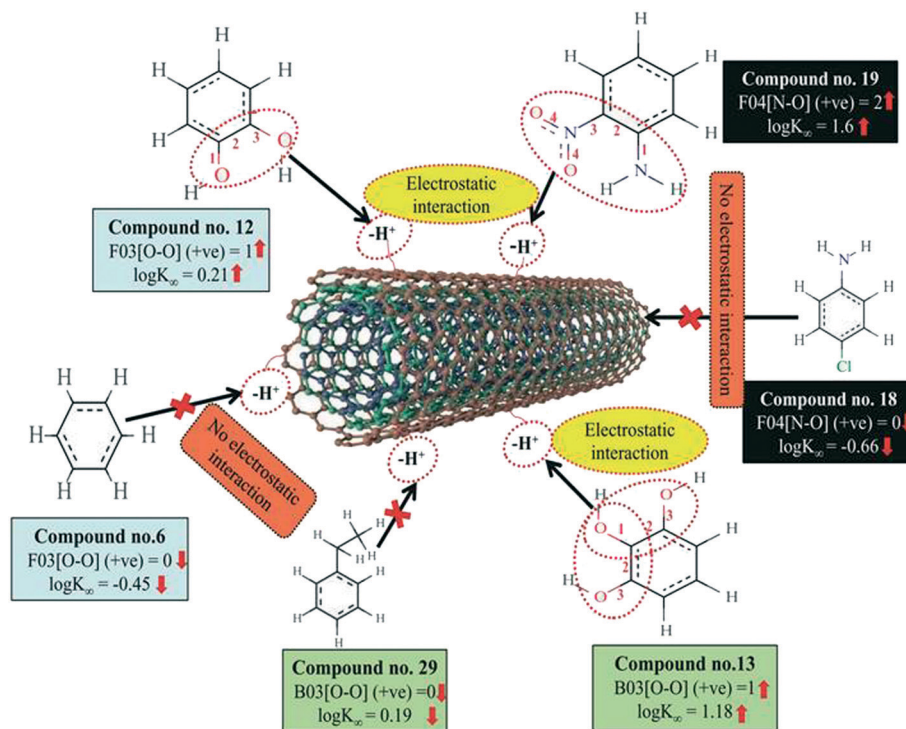


Fig. 6 Mechanistic interpretation of the descriptors related to the electrostatic interactions between organic pollutants and MWCNTs (dataset 1).

electrostatic interactions. The descriptors F03[O-O] and B03[O-O] indicate that the frequency or presence/absence of two electronegative atoms (electron donating group) at the topological distance 3 is essential to enhance the adsorption affinity of organic pollutants to MWCNTs. Thus, there may be a chance to form electrostatic interactions between organic pollutants (negatively charged atom like oxygen atom of the hydroxyl group) and MWCNTs (the sidewall of the CNTs are electrically polarizable and thus polar molecules can easily adhere to their surface). The descriptors involved for electrostatic interactions between organic pollutants and CNTs are represented graphically in Fig. 6.

The 2D atom pair descriptor, B01[C-O], indicates the presence or absence of the C-O bond at topological distance 1. The negative regression coefficient of the descriptor supports that the presence of this fragment at topological distance one is detrimental to the adsorption affinity of organic pollutants by MWCNTs, though it can form hydrogen bonds with MWCNTs. For example, compounds like 1 (pyrene), 57 (2-chloronaphthalene) and 58 (azobenzene) have higher adsorption affinity value due to the absence of such fragments at topological distance 1, whereas compounds like 11 (phenol), 36 (benzyl alcohol) and 42 (phenethyl alcohol) have lower adsorption affinity due to the presence of one C-O bond in each compound.

3.2. Dataset 2: 69 organic pollutants

The significant descriptors obtained from the five MLR models using the adsorption properties ($\log K_{SA}$) of 69 organic pollutants related to the specific surface area of

MWCNTs are Eta_Epsilon_3, X1A, X2A, nOHp, VAdjMat, F04(O-Cl), B05(O-Cl), MLOGP2, T(N...N), O%, and T(O...Cl). We have discussed here all the significant descriptors, which are the key properties for altering the adsorption properties of organic pollutants. The definition, contribution and frequency of the modeled descriptors are shown in Table S5 in the ESI.† The applicability domain of the developed models using the standardization approach showed that one test set compound (compound number 10) for model N1, two test set compounds (compound number 10 and 21) for model N2, one test set compound (compound number 21) for model N3 are situated outside the applicability domain, while in the case of model nos. 4 and 5, all the test set compounds are situated within the domain of applicability. The scatter plot of observed vs. predicted adsorption coefficient related to the specific surface area of MWCNTs for all the MLR models are shown in Fig. 7.

$$\begin{aligned} \text{Model N1. } \log K_{SA} = & 4.29(\pm 2.194) + 0.0965(\pm 0.014) \times \text{O\%} \\ & - 16.4(\pm 4.397) \times \text{X1A} + 0.145(\pm 0.032) \\ & \times \text{T(N}\cdots\text{N)} - 0.0279(\pm 0.009) \\ & \times \text{T(O}\cdots\text{Cl)} - 1.01(\pm 0.294) \\ & \times \text{B05(Cl}\cdots\text{Cl)} + 0.203(\pm 0.022) \\ & \times \text{MLOGP2} \end{aligned}$$

$$n_{\text{training}} = 52, R^2 = 0.845, R^2_{(\text{adj})} = 0.824, Q^2 = 0.798, S = 0.433,$$

$$\text{PRESS} = 11.003, F = 40.79, r^2_{\text{m(LOO)}} = 0.709,$$

$$\Delta r^2_{\text{m(LOO)}} = 0.087, \text{MAE} = \text{Moderate}$$

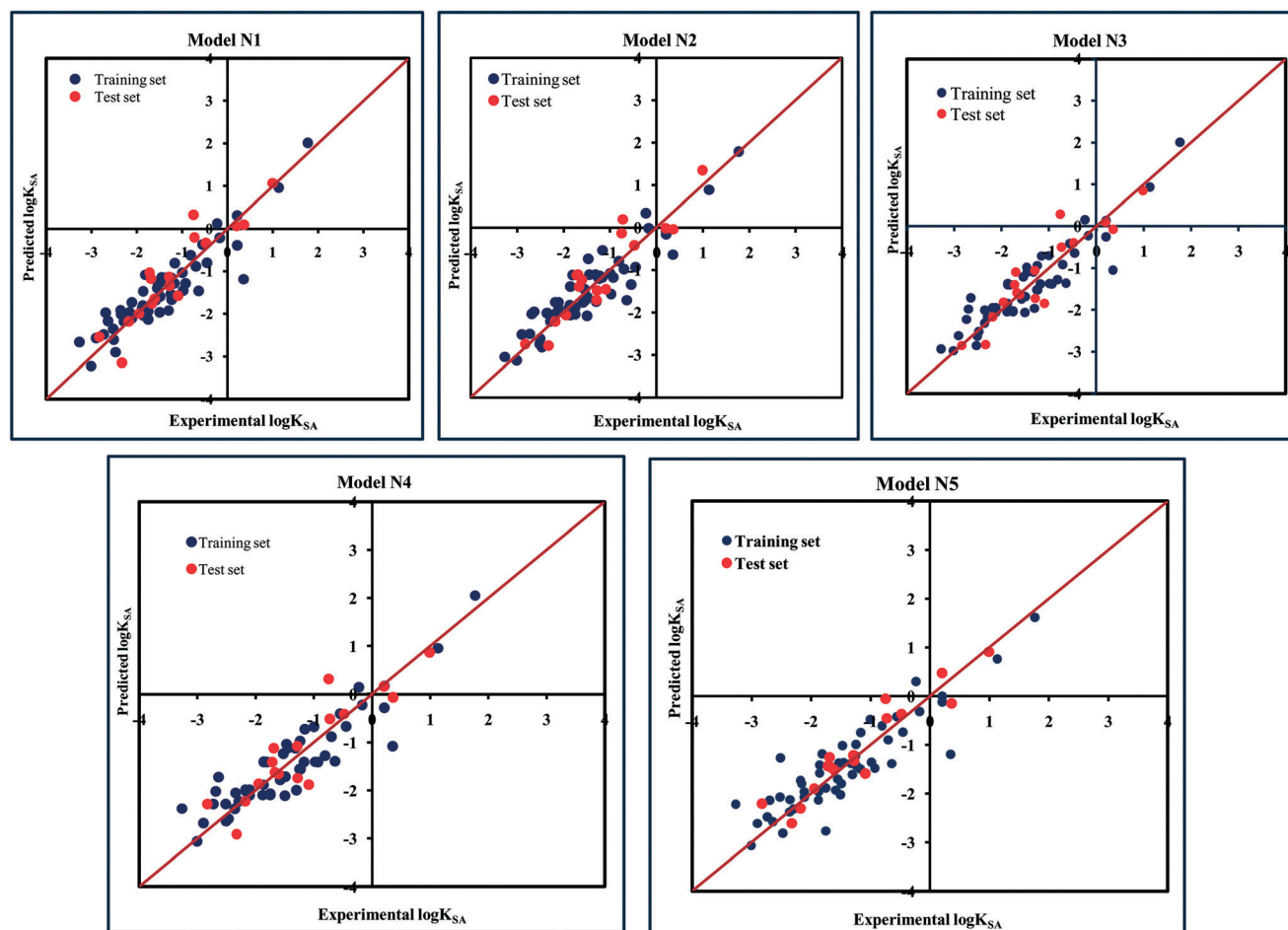
Experimental $\log K_{SA}$ vs predicted $\log K_{SA}$ values of 69 organic pollutants

Fig. 7 The scatter plots of the observed and the predicted adsorption coefficient properties related to the specific surface area of MWCNTs ($\log K_{SA}$) of the developed MLR models (models N1–N5).

$$n_{\text{test}} = 17, Q_{F1}^2 = 0.809, Q_{F2}^2 = 0.795, \overline{r_{m(\text{test})}^2} = 0.783, \Delta r_{m(\text{test})}^2 = 0.048, \\ \text{CCC} = 0.908, \text{MAE} = \text{Moderate}$$

$$n_{\text{test}} = 17, Q_{F1}^2 = 0.830, Q_{F2}^2 = 0.818, \overline{r_{m(\text{test})}^2} = 0.805, \Delta r_{m(\text{test})}^2 = 0.050, \\ \text{CCC} = 0.918, \text{MAE} = \text{Good}$$

$$\text{Model N2. } \log K_{SA} = -7.19(\pm 0.571) + 0.0805(\pm 0.015) \times \text{O\%} \\ - 0.662(\pm 0.323) \times \text{nOHp} \\ - 0.0358(\pm 0.009) \times T(\text{O}\cdots\text{Cl}) \\ - 0.943(\pm 0.294) \times \text{B05}(\text{Cl}\cdots\text{Cl}) \\ + 0.185(\pm 0.019) \times \text{MLOGP2} \\ + 0.958(\pm 0.144) \times \text{VAdjMat}$$

$$\text{Model N3. } \log K_{SA} = -42.3(\pm 7.527) + 0.0973(\pm 0.013) \times \text{O\%} \\ - 0.622(\pm 0.323) \times \text{nOHp} \\ + 0.154(\pm 0.031) \times T(\text{N}\cdots\text{N}) \\ - 0.0407(\pm 0.008) \times T(\text{O}\cdots\text{Cl}) \\ + 0.160(\pm 0.20) \times \text{MLOGP2} \\ + 89.8(\pm 17.51) \times \text{ETA_Epsilon_3}$$

$$n_{\text{training}} = 52, R^2 = 0.842, R_{(\text{adj})}^2 = 0.821, Q^2 = 0.790, S = 0.437, \\ \text{PRESS} = 11.41, F = 39.97, \overline{r_{m(\text{LOO})}^2} = 0.723, \\ \Delta r_{m(\text{LOO})}^2 = 0.114, \text{MAE} = \text{Moderate}$$

$$n_{\text{training}} = 52, R^2 = 0.842, R_{(\text{adj})}^2 = 0.821, Q^2 = 0.788, S = 0.436, \\ \text{PRESS} = 11.512, F = 40.07, \overline{r_{m(\text{LOO})}^2} = 0.714, \\ \Delta r_{m(\text{LOO})}^2 = 0.081, \text{MAE} = \text{Good}$$

$$n_{\text{test}} = 17, Q_{F1}^2 = 0.783, Q_{F2}^2 = 0.768, \overline{r_{m(\text{test})}^2} = 0.712, \Delta r_{m(\text{test})}^2 = 0.14, \\ \text{CCC} = 0.890, \text{MAE} = \text{Good}$$

$$\text{Model N4. } \log K_{\text{SA}} = -42.0(\pm 7.743) + 0.101(\pm 0.014) \times \text{O\%} \\ + 0.159(\pm 0.032) \times T(\text{N}\cdots\text{N}) \\ - 0.0411(\pm 0.008) \times T(\text{O}\cdots\text{Cl}) \\ + 0.168(\pm 0.021) \times \text{MLOGP2} \\ + 88.9(\pm 18.01) \times \text{ETA_Epsilon}_3$$

$$n_{\text{training}} = 52, R^2 = 0.829, R_{(\text{adj})}^2 = 0.811, Q^2 = 0.785, S = 0.449, \\ \text{PRESS} = 11.722, F = 44.73, \overline{r_{m(\text{LOO})}^2} = 0.709, \\ \Delta r_{m(\text{LOO})}^2 = 0.087, \text{MAE} = \text{Good}$$

$$n_{\text{test}} = 17, Q_{F1}^2 = 0.812, Q_{F2}^2 = 0.799, \overline{r_{m(\text{test})}^2} = 0.748, \Delta r_{m(\text{test})}^2 = 0.044, \\ \text{CCC} = 0.903, \text{MAE} = \text{Moderate}$$

$$\text{Model N5. } \log K_{\text{SA}} = 2.49(\pm 1.36) + 0.0757(\pm 0.016) \times \text{O\%} \\ - 17.3(\pm 3.773) \times \text{X2A} + 0.145(\pm 0.036) \\ \times T(\text{N}\cdots\text{N}) - 0.721(\pm 0.144) \\ \times \text{F04}(\text{O}\cdots\text{Cl}) + 0.158(\pm 0.023) \\ \times \text{MLOGP2}$$

$$n_{\text{training}} = 52, R^2 = 0.793, R_{(\text{adj})}^2 = 0.77, Q^2 = 0.743, S = 0.495, \\ \text{PRESS} = 13.955, F = 35.17, \overline{r_{m(\text{LOO})}^2} = 0.709, \\ \Delta r_{m(\text{LOO})}^2 = 0.087, \text{MAE} = \text{Good}$$

$$n_{\text{test}} = 17, Q_{F1}^2 = 0.890, Q_{F2}^2 = 0.882, \overline{r_{m(\text{test})}^2} = 0.836, \Delta r_{m(\text{test})}^2 = 0.090, \\ \text{CCC} = 0.940, \text{MAE} = \text{Good}$$

3.2.1. The descriptors related to the hydrophobic interaction. The descriptor, X1A, indicates an average connectivity index of the order one, it encodes the ‘chi’ value across one bond, which can be calculated on the basis of Kier and Hall’s connectivity index and defined as follows:

$${}^1\text{X} = \sum_{b=1}^B (\delta_i \cdot \delta_j)_b^{-0.5}$$

In this equation, b runs over the 1st order subgraphs having n vertices with B edges; δ_i and δ_j are the number of other vertices attached to vertex i and j , respectively. The negative regression coefficient of this descriptor implies that the higher numerical values of this descriptor are not favorable to enhance the adsorption properties of organic pollutants related to the specific surface area of MWCNTs as shown in compounds 3 (benzene), 56 (ethylbenzene) and 57 (benzyl al-

cohol) (the corresponding numerical values of these compounds are 0.5, 0.491, 0.491, respectively, showing a lower range of adsorption affinity). On the other hand, compounds like 35 (tetracycline), 22 (pyrene) and 26 (phenanthrene) show better adsorption affinity ($\log K_{\text{SA}}$) due to their lower numerical values of this descriptor.

Another significant descriptor, X2A, indicates an average connectivity index of the order 2, and encodes the ‘chi’ value across two bonds, which can be calculated on the basis of Kier and Hall’s connectivity index, defined in the following equation:

$${}^2\text{X} = \sum_{b=2}^B (\delta_i \cdot \delta_j)_b^{-0.5}$$

Here, b runs over the 2nd order subgraphs having n vertices with B edges, δ_i and δ_j are the numbers of other vertices attached to vertex i and j , respectively. This descriptor also has a negative contribution towards the adsorption profile ($\log K_{\text{SA}}$) of organic pollutants by MWCNTs as evidenced by the negative regression coefficient. This indicates that the adsorption properties of organic pollutants decrease with an increase in the numerical value of this descriptor as shown in compounds 3 (benzene), 18 (aniline) and 40 (bromobenzene), and *vice versa* in the case of compounds 22 (pyrene), 26 (phenanthrene) and 35 (tetracycline).

The VAdjMat descriptor represents the vertex adjacency information and gives information about molecular dimension and hydrophobicity. This descriptor can be calculated by using the following formula:

$$\text{VAdjMat} = 1 + \log_2(m)$$

Here, m depicts the number of heavy-heavy bonds. This descriptor contributed positively towards the adsorption properties ($\log K_{\text{SA}}$) of organic pollutants as indicated by the positive regression coefficient. Thus, the higher numerical value of this descriptor is influential toward the adsorption affinity of organic pollutants. This indicates that hydrophobicity plays a crucial role in altering the adsorption properties of organic pollutants by MWCNTs. For example, compounds 22 (pyrene), 26 (phenanthrene) and 35 (tetracycline) show a higher range of adsorption properties as these compounds contain higher numerical values of this descriptor. Compounds 3 (benzene), 55 (iodobenzene) and 46 (chlorobenzene) show a lower range of adsorption properties as these compounds contain higher numerical values of this descriptor. It is therefore suggested that the hydrophobic organic pollutants can easily be adsorbed by MWCNTs through hydrophobic interactions between the pollutants and CNTs.

The next descriptor, MLOGP2, represents the squared Moriguchi octanol–water partition coefficient, calculated from the regression equation of the Moriguchi $\log P$

model^{75,76} consisting of 13 parameters as depicted in the following equation.

$$\begin{aligned} \log P = & -1.244(\text{CX})^{0.6} - 1.017(\text{NO})^{0.9} + 0.406\text{PRX} - 0.145(\text{UB})^{0.8} \\ & + 0.511\text{HB} + 0.268\text{POL} - 2.215\text{AMP} + 0.912\text{ALK} \\ & - 0.392\text{RNG} - 3.684\text{QN} + 0.474\text{NO}_2 + 1.582\text{NCS} \\ & + 0.773\text{BLM} - 1.041 \end{aligned}$$

'CX' depicts the summation of the weighted number of carbon atoms; 'NO' depicts the total number of N and O atoms; 'PRX' represents the proximity effect of N/O; 'UB' represents the number of unsaturated bonds including semi-polar bonds; 'POL' depicts the number of aromatic polar substituents; 'AMP' depicts the amphoteric property; 'ALK' represents the dummy variable for alkanes and alkenes; 'RNG' depicts the indicator variable for the presence of a ring structure, except for benzene and its condensed ring; 'QN' represents quaternary nitrogen; 'NO₂' represents the number of nitro groups; 'HB' represents a dummy variable for the presence of intermolecular hydrogen bonds; 'NCS' depicts isothiocyanato or thiocyanato; 'BLM' represents a dummy variable for the presence of β -lactam.

The positive regression coefficient of this descriptor indicates that hydrophobicity plays a crucial role in regulating the adsorption properties of organic pollutants. The highly hydrophobic organic pollutants can easily be adsorbed by MWCNTs as evidenced by compounds 22 (pyrene), 26 (phenanthrene) and 34 (azobenzene) as their corresponding MLOG2 values are 22.653, 18.762 and 10.539, respectively, whereas hydrophilic molecules are poorly adsorbed by MWCNTs as evidenced by compounds 18 (aniline), 57

(benzylalcohol) and 63 (3-nitroaniline) as their corresponding MLOGP2 values are 2.268, 2.532 and 1.816 respectively. Therefore, it can be inferred that the organic pollutants are adsorbed onto the CNTs through hydrophobic interactions. Thus, for proper adsorption, organic pollutants should be hydrophobic in nature. Note that this was also observed in the case of the VAdjMat descriptor as discussed previously. MLOGP2 is not strictly a 2D descriptor. Here, the term 'intramolecular H-bonds' is used to calculate the MLOGP value, which is conformation dependent.

The information obtained from the descriptors X1A, X2A, VAdjMat and MLOGP2 suggested that the adsorption of organic pollutants related to the specific surface area of MWCNTs may occur through hydrophobic interactions. The molecular connectivity index (X1A and X2A) has a direct relationship with the count of interacting C-H bonds present in a molecule. The number of C-H bonds in a molecule is equal to the number of H atoms. As the C-H bond increases, the hydrophobicity of the molecule increases. The δ value (depends on the number of H atoms, the definition of a δ value for a carbon atom in a molecular graph is: $\delta = 4 - H$) decreases with the average connectivity index. Thus, the hydrophobic interactions between the organic contaminants and MWCNTs are reduced and the adsorption of organic pollutants related to the specific surface area of MWCNTs may also be reduced.⁷⁷

The descriptors VAdjMat and MLOGP2 give information about the hydrophobicity of molecules. It is obvious that the hydrophobic organic pollutants will interact with hydrophobic CNTs through hydrophobic interactions. This implies that the hydrophobic organic pollutants can be easily adsorbed by

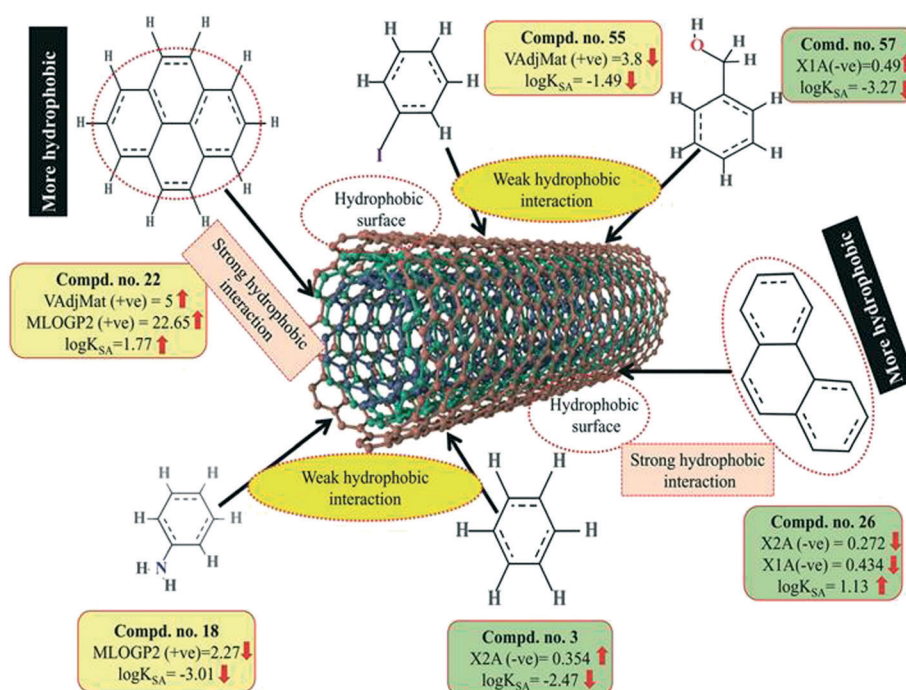


Fig. 8 Mechanistic interpretation of the descriptors related to the hydrophobic interactions between organic pollutants and MWCNTs (dataset 2).

MWCNTs through hydrophobic interactions. The descriptors involved for hydrophobic interaction are graphically depicted in Fig. 8.

3.2.2. The descriptors related to the π - π interactions. A functional group count descriptor, nOHp, describes the number of primary alcohols. The negative regression coefficient of this descriptor points out that the primary alcoholic group is not favored to enhance the adsorption properties ($\log K_{SA}$) of organic pollutants as found in compounds 13 (3-methyl benzyl alcohol) and 57 (benzyl alcohol). On the contrary, organic pollutants that do not contain any primary alcoholic groups have higher adsorption affinities ($\log K_{SA}$) as shown in compounds 22 (pyrene), 26 (phenanthrene) and 34 (azobenzene). Thus, the organic pollutants that do not contain any primary alcoholic groups may be highly adsorbed by MWCNTs.

F04[O-Cl] is a 2D atom pair descriptor that indicates the number of (O-Cl) fragments at a topological distance of 4. The negative regression coefficient of this descriptor indicates that the frequency of the O-Cl fragment at the topological distance 4 is inversely proportional to the adsorption properties of organic pollutants. A higher number for this fragment correlates to lower adsorption properties of organic pollutants, as observed in compounds 7 (dicamba), 61 (3-chlorophenol) and 66 (2,4,5-trichlorophenoxyacetic acid) (these compounds contain 3, 1 and 1 such fragments, respectively, at a topological distance of 4), while a lower numerical value of this descriptor correlates to a higher adsorption property of organic pollutants as observed in compounds 22 (pyrene), 26 (phenanthrene), 34 (azobenzene) and 69 (2,4-dinitrotoluene) (these compounds contain no such fragments at topological distance 4). Thus, the presence of this fragment at the topological distance 4 may hinder the adsorption of the organic pollutants by MWCNTs. The adsorption of organic contaminants to the CNTs decreases when the frequency of the (O-Cl) fragment at topological distance 4 increases. Compound 2 (2,4,6-trichlorophenol) also contains a O-Cl fragment but not at topological distance 4. Therefore, the adsorption affinity related to the specific surface area of the MWCNTs value of compound 2 is ($\log K_{SA}$ value = -0.81) not low as compared to compounds 7 (dicamba), 61 (3-chlorophenol) and 66 (2,4,5-trichlorophenoxyacetic acid) (these compounds contain 3, 1 and 1 such fragments, respectively, at topological distance 4 and the $\log K_{SA}$ values are -2.64, -1.75 and -2.51, respectively).

T(O \cdots Cl), a 2D atom pair descriptor, indicates the sum of the topological distance between oxygen and chlorine. The negative regression coefficient of this descriptor suggests that a higher numerical value of this descriptor is detrimental to enhancing the adsorption properties of organic pollutants related to the specific surface area of MWCNTs as shown in compounds 2 (2,4,6-trichlorophenol), 7 (dicamba) and 66 (2,4,6-trichlorophenoxyacetic acid). On the other hand, the organic pollutants containing no such fragments have higher adsorption properties as shown in compounds 22 (pyrene), 26 (phenanthrene) and 34 (azobenzene). From this observa-

tion, it can be inferred that the organic pollutants without (O \cdots Cl) fragments may be better adsorbed onto the MWCNTs surface.

A 2D atom pair descriptor, B05(Cl-Cl), describes the presence or absence of Cl-Cl fragments at topological distance 5. The negative regression coefficient of this descriptor indicates that the presence of the Cl-Cl fragment at the topological distance 5 may reduce the adsorption property of organic pollutants related to the specific surface area of MWCNTs ($\log K_{SA}$). A higher number of this fragment correlates to lower adsorption property of organic pollutants as observed in compounds 7 (dicamba), 41 (1,2,4-trichlorobenzene) and 66 (2,4,5-trichlorophenoxyacetic acid) (containing one such fragment each) while absence of this fragment in organic pollutants correlates to higher adsorption property as evidenced from compounds 22 (pyrene), 26 (phenanthrene) and 34 (azobenzene). From this descriptor, it can be suggested that the presence of this fragment at topological distance 5 may retard adsorption of the organic pollutants by MWCNTs.

Another 2D atom pair descriptor, T(N \cdots N), indicates the sum of the topological distances between two nitrogen atoms. A positive contribution towards the adsorption properties of organic pollutants related to the specific surface area of MWCNTs ($\log K_{SA}$) indicates that for better adsorption of organic pollutants by MWCNTs, the topological distance between two nitrogen atoms should be greater, as shown in compounds 4 (oxytetracycline), 35 (tetracycline) and 69 (2,4-dinitrotoluene) (as their corresponding topological distances between two nitrogen atoms are 5, 5 and 4, respectively), and *vice versa* in the case of compounds 42 (isophorone), 43 (4-fluorophenol) and 44 (acetophenone). Thus, it can be inferred that the topological distances between two nitrogen atoms should be greater for the better adsorption of organic pollutants by MWCNTs.

As discussed earlier in the introduction section, π - π interactions are one of the key mechanisms for the adsorption of organic pollutants to CNTs. The information obtained from these descriptors, nOHp, F04[O-Cl], B05[Cl-Cl], T(N \cdots N) and T(O \cdots Cl), strongly support this statement. The descriptor nOHp weakens the π - π interaction that occurs between the organic pollutants and CNTs. In this case, the hydroxyl group is alcoholic in nature (aliphatic hydroxyl group) and cannot donate the lone pair of electrons to the aromatic ring (not directly bonded to the aromatic carbon) and ultimately weaken the π - π interactions of the aromatic ring, though it can form hydrogen bonds with the surface modified CNTs. On the other hand, the phenolic hydroxyl group can donate the lone pair of electrons to the aromatic ring (bonded directly to the aromatic carbon atom) as discussed previously (section 3.1), thus strengthening the π - π interactions between organic pollutants and CNTs. In the case of the phenolic hydroxyl group, it can also act as a π donor, but this is not possible in case of the alcoholic hydroxyl group. From this observation, it can be suggested that the aliphatic hydroxyl (alcoholic) group is not favorable for the adsorption affinity of organic pollutants to the CNTs. In case of the descriptors B05[Cl-Cl], T(O \cdots Cl) and

F04[O-Cl], the chlorine atom has an electron inductive effect and decreases the electron density in the benzene ring, which compensates for the electron-donating effect of the oxygen atom (in the case of compounds 7 and 66), even after $-OH$ dissociated into $-O^-$. The withdrawing inductive character of chlorine substituents decreases the electron density of the *p*-chlorophenol ring as compared with that of the phenol ring. Thus, when the O-Cl or Cl-Cl fragment is present in an aromatic molecule, it decreases the electron density of that aromatic ring (as compared with that of the $-OH$ substituted benzene ring (phenolic) or the benzene ring itself) and ultimately, electron donor-acceptor interactions do not occur easily between CNTs and organic contaminants. Hence, the compound could not be easily adsorbed to the MWCNTs. In case of the descriptor T(N \cdots N), the lone pair of electrons of the nitrogen atom can be donated to the ring system (when directly attached) and enhance the π - π interaction with the CNTs. The nitrogen can be present as the amino form (electron donating) or in the nitro form (electron withdrawing). Both forms strengthen the π - π interactions between the organic pollutants and CNTs by increasing or decreasing the π -electron density of the aromatic ring system and act as π electron donor or acceptor, respectively. If the nitrogen is not directly attached to the aromatic ring system, then adsorption happens through electrostatic interactions between the nitrogen of the pollutants and the hydrogen of CNTs by forming dipoles when they are close to each other; the position of the nitrogen atom hardly matters here. The descriptors influencing the π - π interaction are graphically represented in Fig. 9.

3.2.3. The descriptors related to hydrogen bonding interactions. The descriptor, O%, indicates the percentage of oxygen atoms present in a particular molecule. The positive regression coefficient of this descriptor suggests that the presence of oxygen atom is highly influential in the adsorption of the organic pollutants on the surface of MWCNTs. For example, compounds 4 (oxytetracycline), 35 (tetracycline) and 69 (2,4-dinitrotoluene) show better adsorption affinity as their corresponding percentages of oxygen atoms are 15.8, 14.3 and 21.1, respectively. In contrast, compounds 3 (benzene), 18 (aniline) and 24 (4-chloroaniline) show poor adsorption affinity as these compounds do not contain any oxygen atoms. The oxygen atom may be present in different organic pollutants in keto, phenolic (favorable for adsorption) or alcoholic forms (not favorable for adsorption as discussed previously). These different types of oxygen may interact with CNTs in different ways, e.g., hydrogen bonding, strengthening the π - π interactions and electrostatic interactions. On the other hand, a high percentage of oxygen atoms may enhance the polarity of the pollutants. Since the sidewalls of the CNTs are also electrically polarized, the polar group of organic pollutants can easily adhere to the surface of the CNTs. The descriptor involved for hydrogen bonding interactions is given in Fig. 10.

3.2.4. The descriptors related to the electrostatic interactions. The descriptor, Eta_Epsilon_3, indicates the summation of epsilon values relative to the total number of atoms including hydrogen in the connected molecular graph of the reference alkane, which can be calculated by the following equation.

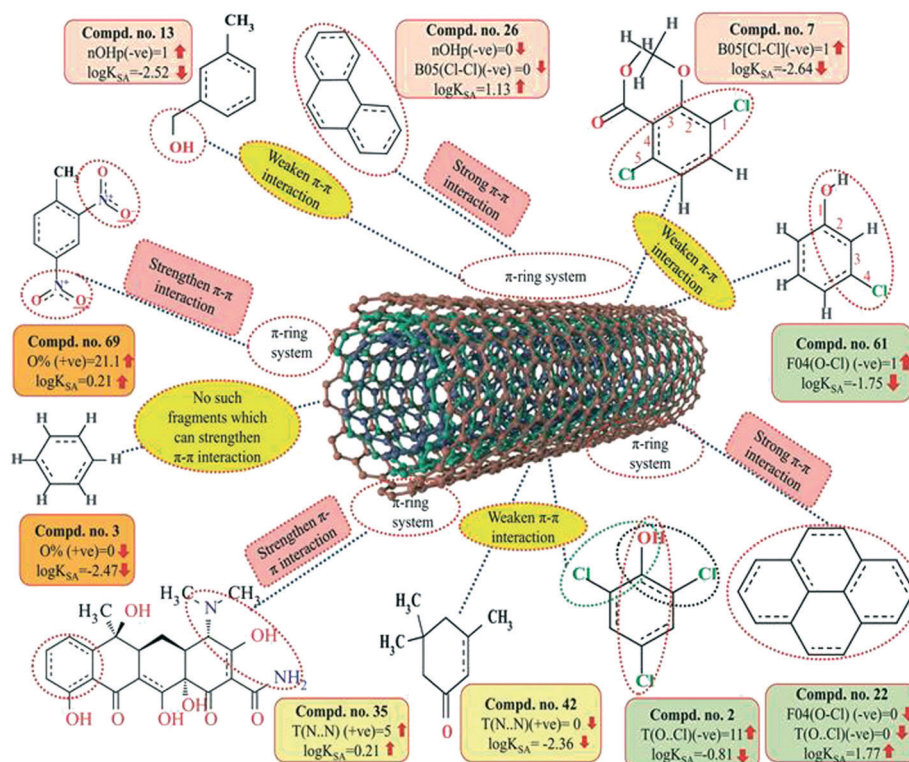


Fig. 9 Mechanistic interpretation of the descriptors related to π - π interactions between organic pollutants and MWCNTs (dataset 2).

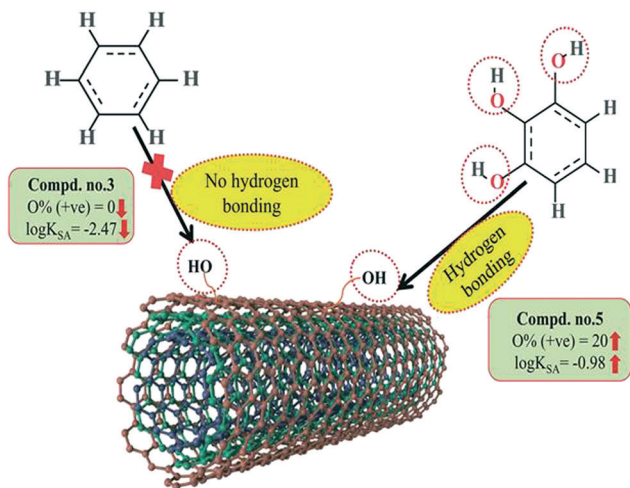


Fig. 10 Mechanistic interpretation of the descriptors related to hydrogen bonding interactions between organic pollutants and MWCNTs (dataset 2).

$$\varepsilon_3 = \varepsilon_R / N_R$$

ε denotes electronegativity, N_R denotes the number of atoms present in the reference alkane. This descriptor has a positive contribution towards the adsorption properties of organic pollutants related to the specific surface area of MWCNTs. This indicates that the electron-rich organic pollutants will be highly adsorbed by MWCNTs. Thus, the higher numerical value (due to strong electrostatic interactions between organic pollutants and CNTs) of this descriptor is required to increase the adsorption properties of organic pollutants by MWCNTs as shown in compounds 22

(pyrene), 26 (phenanthrene) and 35 (tetracycline) and *vice versa* in the case of compounds 7 (dicamba), 13 (3-methylbenzyl alcohol) and 18 (aniline) (due to weak electrostatic interactions between these organic pollutants and CNTs).

The information obtained from the descriptor O% suggests that the organic pollutants can adhere to the surface of MWCNTs by electrostatic interactions. There may be a chance to form electrostatic interactions between organic pollutants (negatively charged atoms like the oxygen atom of the hydroxyl group) and MWCNTs (sidewalls of the CNTs are electrically polarizable, thus polar molecules can easily adhere to their surface). The descriptors involved in electrostatic interactions are shown graphically in Fig. 11.

3.3. Dataset 3 : 29 organic solvents

The significant descriptors obtained from the PLS model using the dispersibility index ($\log C_{\max}$) values of 29 organic solvents to SWCNTs are minsssN, SpMin3_Bhe, VPC-6 and SpMin6_Bhi (arranged according to the variable importance plot, Fig. S2 in ESI[†]). The modeled descriptors, which are the key properties altering the dispersibility indexes of organic solvents, are discussed below. We have also checked the applicability domain of test set compounds using the DModX approach (99% confidence level) to find out whether any test set compounds lie outside of the AD (D-critical = 4.559). The results suggested that the entire test set compounds lie within the AD, except for compound number 29 (Fig. S3 in ESI[†]). The scatter plot of the observed *vs.* predicted dispersibility index of SWCNTs in different solvents are presented in Fig. 12.

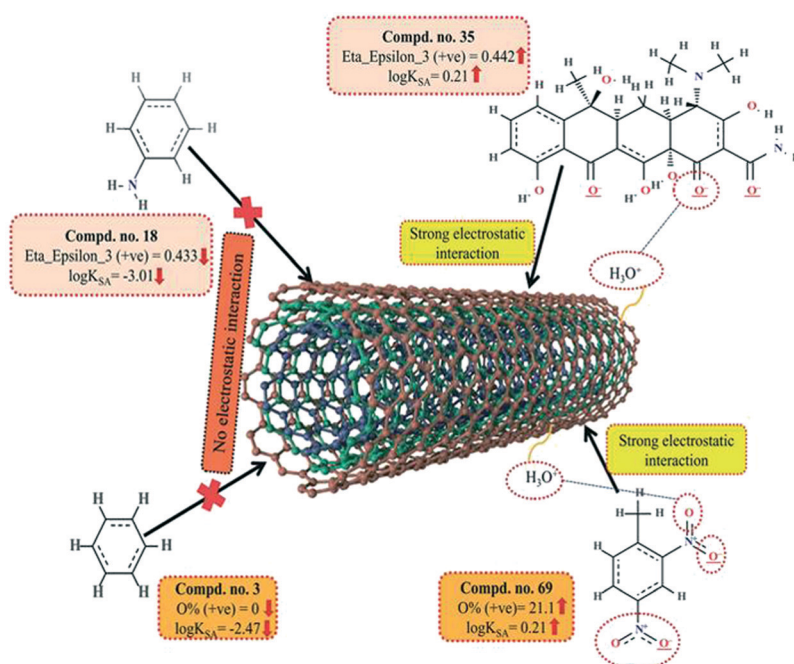


Fig. 11 Mechanistic interpretation of the descriptors related to the electrostatic interactions between organic pollutants and MWCNTs (dataset 2).

$$\text{Model P1. } \log C_{\max} = -1.379 + 1.379 \times \text{VPC-6} - 0.949 \\ \times \text{SpMin3_Bhe} + 0.659 \times \text{minsssN} \\ - 0.375 \times \text{SpMin6_Bhi}$$

$$n_{\text{training}} = 22, R^2 = 0.830, R_{\text{adj}}^2 = 0.810, S = 0.372, F = 29.34,$$

$$\text{PRESS} = 5.164, Q^2 = 0.775, r_{\text{m(LOO)}}^2 = 0.689,$$

$$\Delta r_{\text{m(LOO)}}^2 = 0.115, \text{MAE} = \text{Good},$$

$$n_{\text{test}} = 7, Q_{F1}^2 = 0.945, Q_{F2}^2 = 0.938, r_{\text{m(test)}}^2 = 0.909, \Delta r_{\text{m(test)}}^2 = 0.048,$$

$$\text{CCC} = 0.991, \text{MAE} = \text{Good}$$

The most significant descriptor, minsssN, indicates the minimum atom type E-state >N-. The E-state variable encodes the intrinsic electronic state of each atom present in the molecular graph. The intrinsic electronic state of the atom is changed by the electronic influence of all other atoms in the molecule within the context of the topological character of the molecule. Atoms that possess π and lone pairs of electrons or are terminal atoms possess higher positive values for the E-state index. Atoms that do not have π and lone pairs of electrons and are present at the interior part of a molecule possess lower E-state values. An increase in the minsssN value would indicate the higher electronegativity of the organic solvents, which is beneficial for the dispersibility of SWNTs. The positive regression coefficient of this descriptor indicates that nitrogen atoms connected to other heavy atoms play an important role in influencing the dispersibility of SWNTs in different organic solvents. The numerical values of this descriptor are directly proportional to the dispersibility of SWNTs, suggesting that the dispersibility index of the SWNTs will increase with increasing the number of such fragments as evidenced by the compounds 1 (1,3-dimethyltetrahydro-2(1H)-pyrimidinone), 2 (1-butylpyrrolidin-2-

one) and 5 (3-(2-oxo-1-pyrrolidinyl)propanenitrile). On the other hand, the absence of such fragments in different organic solvents decreases the dispersibility index of SWCNTs as shown in compounds 24 (cyclohexanone), 27 (formamide) and 28 (benzyl alcohol). Thus, from this descriptor, it can be suggested that the dispersibility of CNTs may be enhanced through electrostatic interactions.

The second highest significant descriptor, *SpMin3_Bhe*, is defined as the smallest absolute eigenvalue of Burden modified matrix-n3/weighted by the relative Sanderson electronegativities.⁷⁸ The negative contribution shown by *SpMin3_Bhe* indicates that the dispersibility index of SWCNTs in various solvents can be increased by decreasing the numerical value of *SpMin3_Bhe* as shown in compounds 9 (dimethylimidazolidinone), 10 (dimethyl acetamide) and 16 (acrylic acid). On the other hand, the dispersibility of SWCNTs can be decreased by increasing the numerical value of *SpMin3_Bhe* as shown in compounds 22 (benzyl benzoate) and 26 (triethyleneglycol). The *SpMin3_Bhe* descriptor weighted by the relative Sanderson electronegativity suggests that the electronegativity of the solvents and polar interactions with CNTs play an important role in the dispersibility of the SWCNTs. It can be concluded that polar interactions can have an optimum value. Thus, polar solvents with low donor number are preferred for the dispersibility of the CNTs or it would be better to state that solvents with medium polarity are satisfactory.

The third highest significant descriptor, VPC-6, is a type of topological descriptor, which indicates the chi valance path cluster of order 6. This descriptor differentiates the molecules according to their size, degree of branching, flexibility and overall shape. Chi cluster descriptor (VPC-6) is an indicator of the *n*th degree of branching and thus implicates the effect of substitution in a molecule. The organic solvent molecules that are relatively compact have higher values of this descriptor,⁷⁹ suggesting that a small sized molecule with compactness is most probably a better solvent for SWCNTs. It has a positive contribution toward the dispersibility index of SWCNTs in different organic solvents. This indicates that the degree of branching of organic solvents increases the dispersibility index of SWCNTs as shown in compounds 1 (1,3-dimethyltetrahydro-2(1H)-pyrimidinone), 3 (1-benzylpyrrolidin-2-one), and 9 (dimethyl-imidazolidinone), and *vice versa* in case of compounds 10 (dimethyl acetamide), 16 (acrylic acid) and 17 (2,2'-thiodiethanol).

The least significant descriptor, SpMin6_Bhi indicates the smallest absolute eigenvalue of Burden modified matrix - n6/weighted by the relative first ionization potential.

A modified Burden matrix Q is defined as follows:

$$[Q]_{ij} = Z_i + 0.1\delta_i + 0.01 \times n_i^\pi \text{ and } [Q]_{ij} = 0.4/d_{ij}$$

where, Z_i depicts the atomic number of the *i*th atom, d_i depicts the number of non-hydrogen neighbors of the *i*th atom (*i.e.*, the vertex degree), n_i^π depicts the number of π electrons, and d_{ij} depicts the topological distance between the *i*th and

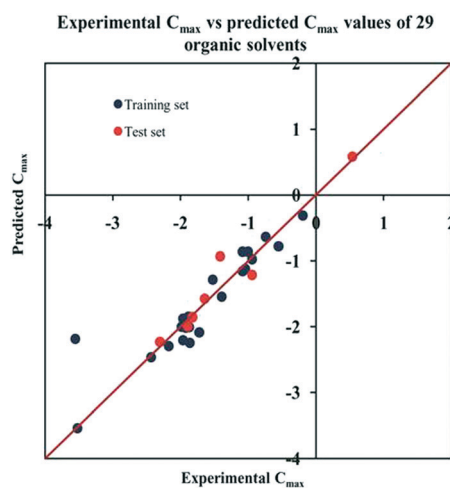


Fig. 12 The scatter plot of the observed and the predicted dispersibility index of SWCNTs ($\log C_{\max}$) of the developed PLS model (model P1).

*j*th atoms.⁷⁸ A larger ionization potential of a molecule suggests that higher energy is required to convert the molecule into cationic form, whereas a smaller ionization potential can easily convert the molecule into cationic form, which helps in the easy interaction of the cationic form of the molecule to the π -system of the carbon nanotube through π -cationic interactions. This descriptor is inversely proportional to the dispersibility of SWNTs, suggesting that with increasing the ionization potential, the dispersibility index of the SWNTs decreases as evidenced by compounds 27 (formamide), 16 (acrylic acid), and 9 (dimethyl-imidazolidinone). On the other hand, the dispersibility index of organic solvents increases in the case of compounds 2 (1-butylpyrrolidin-2-one) and 5 [3-(2-oxo-1-pyrrolidinyl)propanenitrile]. The effects of the contributed descriptors on the dispersibility of SWCNTs in diverse organic solvents are summarized graphically in Fig. 13.

4. Overview and conclusions

MLR and PLS regression-based strategies were employed to develop QSPR models of organic pollutants (datasets 1 & 2) and organic solvents (dataset 3). Multiple endpoints related to CNTs (adsorption coefficient, adsorption coefficient related to specific surface area of MWCNTs and dispersibility index) were used to explore the key structural features that influence the adsorption and dispersibility of the investigated molecules towards MWCNTs and SWCNTs, respectively. The models were developed using 2D descriptors only. Prior to the development of the final models, different strategies for variable selection were performed to extract the most significant descriptors for the generation of the final MLR (5

models for both datasets 1 and 2) and PLS (a single model for dataset 3) models. Extensive validation of the developed models was performed, which showed good predictability and robustness. The QSPR models were developed in compliance with the OCED principles. We also used the “Intelligent consensus predictor” tool to explore whether the quality of the predictions of test set compounds could be enhanced through an “intelligent” selection of multiple MLR models (in the case of datasets 1 and 2). The results showed that based on the MAE-based criteria, the consensus predictions of multiple MLR models are better than the results obtained from the individual models. In both cases, the winning model was CM3. The insights obtained from the developed MLR models for datasets 1 and 2 are as follows: (i) the descriptors like U_i , F03[O–O], F04[N–O], ETA_BetaP, nOHp, O%, T(N··N), T(O··Cl) and F04[O–Cl] influence the adsorption of organic pollutants either by π - π interactions or by strengthening π - π interactions. (ii) $nArOH$, F03[O–O], B03[O–O], nHBint, F04[N–O], Eta_Epsilon_3 and O% descriptors favor the adsorption of organic pollutants through electrostatic interactions. (iii) The organic pollutants adsorbed through hydrogen bonding interactions are indicated by $nArOH$, F03[O–O], B03[O–O], nHBint, F04[N–O] and O%. (iv) The descriptors minsCH₃, B06[C–Cl], X0v, VAdjMat, MLOGP2, X2A and X1A are essential for the adsorption of organic pollutants through hydrophobic interactions. These observations were further supported by the following discussion: the organic adsorbates of CNTs were mostly aromatic compounds, confirming that aromatic compounds have a better interaction with CNTs than the non-aromatic pollutants, due to their π electron richness and flat conformation. The

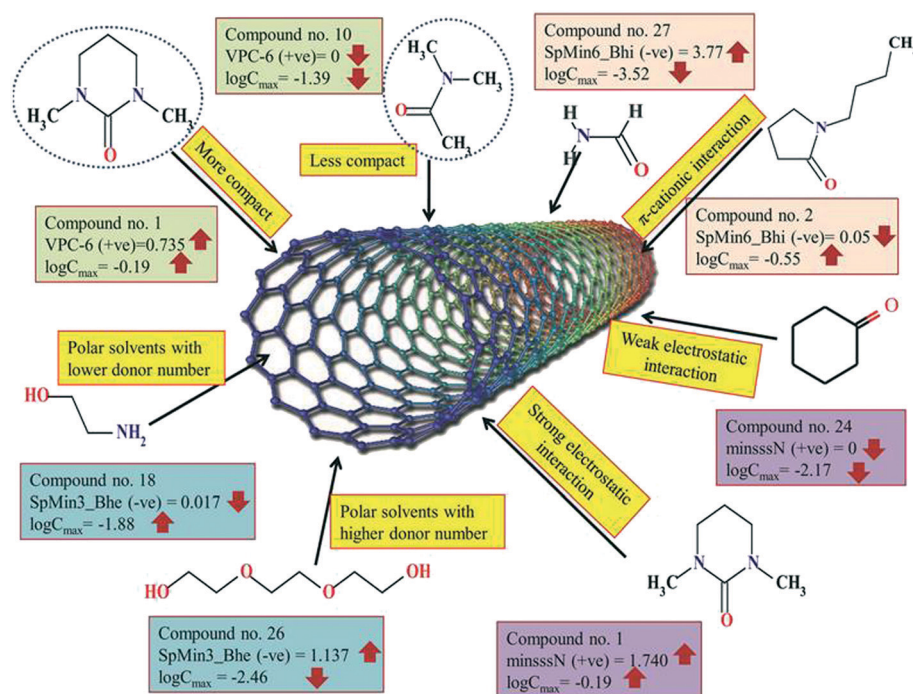


Fig. 13 The effects of the contributed descriptors on the dispersibility of SWCNTs in diverse organic solvents.

systematic understanding of aromatic contaminants is therefore critical since aromaticity plays an important role in adsorption. Several studies have suggested that π - π interactions are crucial for the adsorption of organic compounds to CNTs,^{71,80,81} which in turn depends on the size and shape of the molecules, due to the curvature of the CNTs and its substituents. The π -system of the organic pollutants interacts with the π -system of the CNTs through π - π interactions and the interactions increase with the number of aromatic rings in the adsorbates.^{39,82} Both electron withdrawing groups (e.g. $-\text{NO}_2$ and $-\text{Cl}$) and electron donating groups (e.g. $-\text{NH}_2$, $-\text{OH}$) strengthen the π - π interactions between the pollutants and MWCNTs^{73,74} by acting as π -electron acceptors and π -electron donors, respectively. The hydroxyl group was investigated as an electron donating substituent on adsorptive interactions among pollutants and MWCNTs, since the hydroxyls, by dissociating to $-\text{O}^-$ (which has stronger electron donating ability), strengthen the n- π electron donor-acceptor (EDA) mechanism. Compounds with no aromatic ring (no π electrons) interact through hydrophobic forces. A study also suggested that CNTs act as strong adsorbents for hydrophobic compounds due to hydrophobic interactions.^{15,16,33,83-85} Hydroxyl groups (phenolic form) can interact through various means, such as (i) hydrophobic interactions (ii) electrostatic interactions (both attraction and repulsion) (iii) hydrogen bonding interactions and (iv) enhancing π - π interactions. As the number of hydroxyl groups (phenolics) in the pollutants increases, the hydrophobicity decreases. Thus, it can be considered as a major factor in the adsorption of phenolics to CNTs. Hydrogen bonding can also be a major interaction between hydroxyl-containing pollutants and substituted carbon nanotubes.^{86,87} Hydroxyl and amino group interactions can be related to the electronic features. In one experiment, it was observed that 1-naphthylamine has better adsorption to treated CNTs than the untreated CNTs, and there was an additional observation that although both 2,4-dichlorophenol and 2-naphthol contain an $-\text{OH}$ group, the adsorption of 2-naphthol was more significant with variation in the functionality of CNTs.⁸⁸ This indicates that when the adsorbates possess electronic properties, the functionality of nanotubes helps with the improvement of adsorption.⁸⁸ Chen *et al.*⁸⁹ reported that nitro group containing pollutants show stronger adsorption than non-polar aromatics. This indicates that along with hydrophobic interactions, there is some other essential interaction that controls the adsorption, which is comparable to the π -electron polarizability that is related to aromatic compounds and electron donating as well as accepting properties, similar to compounds having more than two nitro groups. Nitroaromatic compounds, besides being polar in nature, have electron accepting capacity when interacting with adsorbents having high electron polarizability properties and also have high electron conjugation with the π -electrons of CNTs. Thus, the higher affinity of nitro aromatic compounds as compared to other pollutants is due to π - π electron donor-acceptor interactions; since nitrogen is a strong electron-withdrawing atom, it acts as a π -acceptor and

carbon nanotubes act as the π -donor.⁹⁰⁻⁹³ Hydrogen bonding is also possible between nitro groups of the pollutants, which act as H-acceptors and functional group-substituted carbon nanotubes. The presence of two chlorine atoms causes the electron inductive effect, which may cause a reduction in the electron density of the aromatic ring attached to it, as suggested by Sulaymon and Ahmed *et al.*;⁹⁴ the electron donating effect of the hydroxyl atom attached to the aromatic ring compensates for this by dissociating into the stronger electron donor like $-\text{O}^-$ (oxygen). We can, therefore, conclude that the adsorption of the organic pollutants to the CNTs can be enhanced by the following: a greater number of aromatic rings, high unsaturation or electron richness of the molecule, the presence of polar groups substituted on the aromatic ring, the presence of two oxygen atoms at a topological distance of 3, the presence of nitrogen and oxygen atoms at the topological distance of 4, the size of the molecules, and the hydrophobic surface of the molecules. On the other hand, the presence of carbon and oxygen atoms at a topological distance of 1, aliphatic primary alcohols, the presence of two chlorine atoms at topological distance 5 and the presence of oxygen and chlorine atoms at topological distance 4 may be detrimental and can retard the adsorption of organic pollutants. From the insights obtained from the PLS model for dataset 3, we have interpreted that the organic solvents bearing the $>\text{N}$ - fragment, polar solvents with low donor number, compact molecules and lower ionization potential may be better solvents to enhance the dispersibility of SWCNTs. Dispersibility is directly correlated to the adsorption properties of molecules to CNTs. This PLS model and contributed descriptors can help with the understanding of the mechanism of the dispersion process and predict organic solvents that improve the dispersibility of SWCNTs and may overcome the drawbacks of SWCNTs. This work may, therefore, be helpful in the removal of the harmful and toxic contaminants/disposal of the by-products from the various industries, making it possible to achieve a pollution-free environment.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

Financial assistance from the AICTE, New Delhi in the form of a fellowship to JR and SG is thankfully acknowledged. PKO thanks the UGC, New Delhi for financial assistance in the form of a fellowship (Letter number and date: F./PDFSS-2015-17-WES-11996; dated: 06/04/2016). KR thanks CSIR, New Delhi for financial assistance under a Major Research project (CSIR Project No. 01(2895)/17/EMR-II).

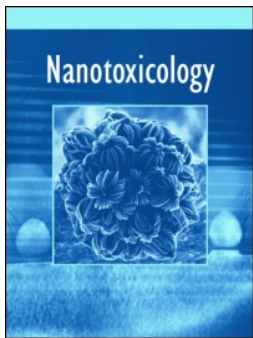
References

- 1 U. K. Garg, M. P. Kaur, V. K. Garg and D. Sud, Removal of hexavalent Cr from aqueous 19 solutions by agricultural waste biomass, *J. Hazard. Mater.*, 2007, **140**, 60-68.

- 2 J. M. Randall, E. Hautala and A. C. Waiss Jr, Removal and recycling of heavy metal ions from mining and industrial waste streams with agricultural by-products, in *Proceedings of the fourth mineral waste utilization symposium*, Chicago, 1974.
- 3 D. J. Ferner, Toxicity, heavy metals, *Med. J.*, 2001, 2(5), 1.
- 4 Y. Lu, S. Song, R. Wang, Z. Liu, J. Meng, A. J. Sweetman, A. Jenkins, R. C. Ferrier, H. Li, W. Luo and T. Wang, Impacts of soil and water pollution on food safety and health risks in China, *Environ. Int.*, 2015, 77, 5–15.
- 5 L. B. Franklin, *Wastewater engineering: treatment, disposal and reuse*, McGraw Hill, New York, 1991.
- 6 R. L. Droste, *Theory and practice of water and wastewater treatment*, Wiley, New York, 1997.
- 7 R. N. Goyal, V. K. Gupta, A. Sangal and N. Bachheti, Voltammetric determination of uric acid at a fullerene-C60-modified glassy carbon electrode, *Electroanalysis*, 2005, 17(24), 2217–2223.
- 8 R. N. Goyal, V. K. Gupta and N. Bachheti, Voltammetric determination of adenosine and guanosine using fullerene-C60-modified glassy carbon electrode, *Talanta*, 2007, 71(3), 1110–1117.
- 9 R. N. Goyal, V. K. Gupta and N. Bachheti, Fullerene-C60-modified electrode as a sensitive voltammetric sensor for detection of nandrolone, *Anal. Chim. Acta*, 2007, 597, 82–89.
- 10 R. N. Goyal, V. K. Gupta, N. Bachheti and R. A. Sharma, Electrochemical sensor for the determination of dopamine in presence of high concentration of ascorbic acid using a fullerene-C60 coated gold electrode, *Electroanalysis*, 2008, 20, 757–764.
- 11 R. N. Goyal, M. Oyama, V. K. Gupta, S. P. Singh and S. Chatterjee, Sensors for 5-hydroxytryptamine and 5-hydroxyindole acetic acid based on nanomaterial modified electrodes, *Sens. Actuators, B*, 2008, 134, 816–821.
- 12 R. N. Goyal, V. K. Gupta and S. Chatterjee, Fullerene-C60-modified edge plane pyrolytic graphite electrode for the determination of dexamethasone in pharmaceutical formulations and human biological fluids, *Biosens. Bioelectron.*, 2009, 24, 1649–1654.
- 13 D. Z. John, *Handbook of drinking water quality: standards and controls*, Van Nostrand Reinhold, New York, 1990.
- 14 E. A. Laws, *Aquatic pollution: an introductory text*, Wiley, New York, 3rd edn, 2000.
- 15 K. Yang, L. Z. Zhu and B. S. Xing, Adsorption of polycyclic aromatic hydrocarbons by carbon nanomaterials, *Environ. Sci. Technol.*, 2006, 40, 1855–1861.
- 16 K. Yang, X. Wang, L. Zhu and B. Xing, Competitive sorption of pyrene, phenanthrene, and naphthalene on multiwalled carbon nanotubes, *Environ. Sci. Technol.*, 2006, 40, 5804–5810.
- 17 Y. H. Li, Z. Di, J. Ding, D. Wu, Z. Luan and Y. Zhu, Adsorption thermodynamic, kinetic and desorption studies of Pb²⁺ on carbon nanotubes, *Water Res.*, 2005, 39(4), 605–609.
- 18 H. M. Al-Saidi, M. A. Abdel-Fadeel, A. Z. El-Sonbati and A. A. El-Bindary, Multi-walled carbon nanotubes as an adsorbent material for the solid phase extraction of bismuth from aqueous media: kinetic and thermodynamic studies and analytical applications, *J. Mol. Liq.*, 2016, 216, 693–698.
- 19 S. Kumar, G. Bhanjana, N. Dilbaghi and A. Umar, Multi walled carbon nanotubes as sorbent for removal of crystal violet, *J. Nanosci. Nanotechnol.*, 2014, 14, 7054–7059.
- 20 S. Mosayebidorcheh and M. Hatami, Heat transfer analysis in carbon nanotube-water between rotating disks under thermal radiation conditions, *J. Mol. Liq.*, 2017, 240, 258–267.
- 21 N. Nakashima, Soluble carbon nanotubes: Fundamental and applications, *Int. J. Nanosci.*, 2005, 4, 119–137.
- 22 D. A. Britz and A. N. Khlobystov, Noncovalent interactions of molecules with single walled carbon nanotubes, *Chem. Soc. Rev.*, 2006, 35(7), 637–659.
- 23 L. A. Girifalco, M. Hodak and R. S. Lee, Carbon nanotubes, buckyballs, ropes, and a universal graphitic potential, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2000, 62(19), 13104.
- 24 E. Hammel, X. Tang, M. Trampert, T. Schmitt, K. Mauthner, A. Eder and P. Pötschke, Carbon nanofibers for composite applications, *Carbon*, 2004, 42(5–6), 1153–1158.
- 25 T. Liu, I. Y. Phang, L. Shen, S. Y. Chow and W. D. Zhang, Morphology and mechanical properties of multiwalled carbon nanotubes reinforced nylon-6 composites, *Macromolecules*, 2004, 37(19), 7214–7222.
- 26 Y. S. Song and J. R. Youn, Influence of dispersion states of carbon nanotubes on physical properties of epoxy nanocomposites, *Carbon*, 2005, 43(7), 1378–1385.
- 27 K. E. Geckeler and T. Premkumar, Carbon nanotubes: are they dispersed or dissolved in liquids?, *Nanoscale Res. Lett.*, 2011, 6(1), 136.
- 28 A. Abbas, A. M. Al-Amer, T. Laoui, M. J. Al-Marri, M. S. Nasser, M. Khraisheh and M. A. Atieh, Heavy metal removal from aqueous solution by advanced carbon nanotubes: critical review of adsorption applications, *Sep. Purif. Technol.*, 2016, 157, 141–161.
- 29 O. V. Kharissova, B. I. Kharisov and E. G. de Casas Ortiz, Dispersion of carbon nanotubes in water and non-aqueous solvents, *RSC Adv.*, 2013, 3(47), 24812–24852.
- 30 H. Hyung, J. D. Fortner, J. B. Hughes and J. H. Kim, Natural organic matter stabilizes carbon nanotubes in the aqueous phase, *Environ. Sci. Technol.*, 2007, 41, 179–184.
- 31 R. Q. Long and R. T. Yang, Carbon nanotubes as superior sorbent for dioxin removal, *J. Am. Chem. Soc.*, 2001, 123, 2058–2059.
- 32 G. P. Rao, C. Lu and F. Su, Sorption of divalent heavy metal ions from aqueous solution by carbon nanotubes: a review, *Sep. Purif. Technol.*, 2007, 58, 224–231.
- 33 J. Hilding, E. A. Grulke, S. B. Sinnott, D. Qian, R. Andrews and M. Jagtoyen, Sorption of butane on carbon multiwall nanotubes at room temperature, *Langmuir*, 2001, 17, 7540–7544.
- 34 C. S. Lu, Y. L. Chung and K. F. Chang, Adsorption of trihalomethanes from water with carbon nanotubes, *Water Res.*, 2005, 39, 1183–1189.

- 35 C. J. M. Chin, L. C. Shih, H. J. Tsai and T. K. Liu, Adsorption of o-xylene and p-xylene from water by SWCNTs, *Carbon*, 2007, **45**, 1254–1260.
- 36 Q. Liao, J. Sun and L. Gao, Adsorption of chlorophenols by multiwalled carbon nanotubes treated with HNO₃ and NH₃, *Carbon*, 2008, **46**, 553–555.
- 37 X. J. Peng, Y. H. Li, Z. K. Luan, Z. C. Di, H. Y. Wang, B. H. Tian and Z. P. Jia, Adsorption of 1, 2-dichlorobenzene from water to carbon nanotubes, *Chem. Phys. Lett.*, 2003, **376**, 154–158.
- 38 Q. Liao, J. Sun and L. Gao, The adsorption of resorcinol from water using multi-walled carbon nanotubes, *Colloids Surf.*, 2008, **312**, 160–165.
- 39 S. Gotovac, H. Honda, Y. Hattori, K. Takahashi, H. Kanoh and K. Kaneko, Effect of nanoscale curvature of single-walled carbon nanotubes on adsorption of polycyclic aromatic hydrocarbons, *Nano Lett.*, 2007, **7**, 583–587.
- 40 D. C. Luehrs, J. P. Hickey, P. E. Nilsen, K. A. Godbole and T. N. Rogers, Linear solvation energy relationship of the limiting partition coefficient of organic solutes between water and activated carbon, *Environ. Sci. Technol.*, 1996, **30**, 143–152.
- 41 O. G. Apul, Q. Wang, T. Shao, J. R. Rieck and T. Karanfi, Predictive model development for adsorption of aromatic contaminants by multi-walled carbon nanotubes, *Environ. Sci. Technol.*, 2012, **47**, 2295–2303.
- 42 M. Rahimi-Nasrabadi, R. Akhoondi, S. M. Pourmortazavi and F. Ahmadi, Predicting adsorption of aromatic compounds by carbon nanotubes based on quantitative structure property relationship principles, *J. Mol. Struct.*, 2015, **1099**, 510–515.
- 43 X. R. Xia, N. A. Monteiro-Riviere and J. E. Riviere, An index for characterization of nanomaterials in biological systems, *Nat. Nanotechnol.*, 2010, **5**, 671–675.
- 44 V. Chayawan, Quantum-mechanical parameters for the risk assessment of multiwalled carbon-nanotubes: a study using adsorption of probe compounds and its application to biomolecules, *Environ. Pollut.*, 2016, **218**, 615–624.
- 45 O. G. Apul, P. Xuan, F. Luo and T. Karanfil, Development of a 3D QSPR model for adsorption of aromatic compounds by carbon nanotubes: comparison of multiple linear regression, artificial neural network and support vector machine, *RSC Adv.*, 2013, **3**, 23924–23934.
- 46 O. G. Apul, Y. Zhou and T. Karanfil, Mechanisms and modeling of halogenated aliphatic contaminant adsorption by carbon nanotubes, *J. Hazard. Mater.*, 2015, **295**, 138–144.
- 47 Z. Hassanzadeh, M. Kompany-Zareh, R. Ghavami, S. Gholami and A. Malek-Khatabi, Combining radial basis function neural network with genetic algorithm to QSPR modeling of adsorption on multi-walled carbon nanotubes surface, *J. Mol. Struct.*, 2015, **1098**, 191–198.
- 48 H. Yilmaz, B. Rasulev and J. Leszczynski, Modeling the dispersibility of single walled carbon nanotubes in organic solvents by quantitative structure-activity relationship approach, *Nanomaterials*, 2015, **5**, 778–791.
- 49 M. Salahinejad and E. Zolfonoun, QSAR studies of the dispersion of SWNTs in different organic solvents, *J. Nanopart. Res.*, 2013, **15**, 2028.
- 50 M. Rofouei, M. Salahinejad and J. B. Ghasemi, An alignment independent 3D-QSAR modeling of dispersibility of single-walled carbon nanotubes in different organic solvents, *Fullerenes, Nanotubes, Carbon Nanostruct.*, 2014, **22**, 605–617.
- 51 A. Heidari and M. H. Fatemi, Hybrid docking-Nano-QSPR: an alternative approach for prediction of chemicals adsorption on nanoparticles, *NANO*, 2016, **11**, 1650078.
- 52 S. D. Bergin, Z. Sun, D. Rickard, P. V. Streich, J. P. Hamilton and J. N. Coleman, Multicomponent solubility parameters for single-walled carbon, nanotube–solvent mixtures, *ACS Nano*, 2009, **3**, 2340–2350.
- 53 <http://www.chemaxon.com>.
- 54 http://www.taletе.mi.it/products/dragon_description.htm.
- 55 <http://www.yapcsoft.com/dd/padeldescriptor>.
- 56 S. Das, P. K. Ojha and K. Roy, Multilayered variable selection in QSPR: a case study of modeling melting point of bromide ionic liquids, *Int. J. Quant. Struct.-Prop. Relat.*, 2017, **2(1)**, 106–124.
- 57 P. K. Ojha and K. Roy, Comparative QSARs for antimalarial endochins: importance of descriptor-thinning and noise reduction prior to feature selection, *Chemom. Intell. Lab. Syst.*, 2011, **109(2)**, 146–161.
- 58 http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab.
- 59 K. Roy, R. N. Das, P. Ambure and R. B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, *Chemom. Intell. Lab. Syst.*, 2016, **152**, 18–33.
- 60 K. Roy, P. Ambure, S. Kar and P. K. Ojha, Is it possible to improve the quality of predictions from an “intelligent” use of multiple QSAR/QSPR/QSTR models?, *J. Chemom.*, 2018, **32(4)**, 2992.
- 61 R. B. Darlington, in *Regression and linear models*, New York, McGraw-Hill, 1990.
- 62 D. Rogers and A. J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 854–866.
- 63 P. K. Ojha, I. Mira, R. N. Das and K. Roy, Further exploring r_m^2 metrics for validation of QSPR models, *Chemom. Intell. Lab. Syst.*, 2011, **107(1)**, 194–205.
- 64 I. Lawrence and K. Lin, Assay validation using the concordance correlation coefficient, *Biometrics*, 1992, 599–604.
- 65 N. Chirico and P. Gramatica, Real external predictivity of QSAR models: how to evaluate it? Comparison of different validation criteria and proposal of using the concordance correlation coefficient, *J. Chem. Inf. Model.*, 2011, **51(9)**, 2320–2335.
- 66 K. Roy, S. Kar and P. Ambure, On a simple approach for determining applicability domain of QSAR models, *Chemom. Intell. Lab. Syst.*, 2015, **145**, 22–29.
- 67 S. Wold, M. Sjöström and L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.
- 68 UMETRICS, *UMETRICS SIMCA-P 10.0*, Umea, Sweden, 2002, info@umetrics.com, www.umetrics.com.
- 69 <http://www.miniTab.com/en-US/default.aspx>.
- 70 SPSS is statistical software of SPSS Inc., USA, 1999.

- 71 Y. Zhang, S. L. Yuan, W. W. Zhou, J. J. Xu and Y. Li, Spectroscopic evidence and molecular simulation investigation of the pi-pi interaction between pyrene molecules and carbon nanotubes, *J. Nanosci. Nanotechnol.*, 2007, 7, 2366–2375.
- 72 L. H. Hall and L. B. Kier, Electrotological state indices for atom types: a novel combination of electronic, topological, and valence state information, *J. Chem. Inf. Comput. Sci.*, 1995, 35(6), 1039–1045.
- 73 L. M. Woods, S. C. Bădescu and T. L. Reinecke, Adsorption of simple benzene derivatives on carbon nanotubes, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, 75(1–9), 155415.
- 74 A. Star, T. R. Han, J. C. P. Gabriel, K. Bradley and G. Gruner, Interaction of aromatic compounds with carbon nanotubes: correlation to the Hammett parameter of the substituent and measured carbon nanotube FET response, *Nano Lett.*, 2003, 3, 1421–1423.
- 75 I. Moriguchi, S. Hirono, I. Nakagome and H. Hirano, Comparison of reliability of log P values for drugs calculated by several methods, *Chem. Pharm. Bull.*, 1994, 42(4), 976–978.
- 76 P. K. Ojha and K. Roy, Development of a robust and validated 2D-QSPR model for sweetness potency of diverse functional organic molecules, *Food Chem. Toxicol.*, 2017, 112, 551–562.
- 77 L. B. Kier and L. H. Hall, The meaning of molecular connectivity: A bimolecular accessibility model, *Croat. Chem. Acta*, 2002, 75(2), 371–382.
- 78 R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics: volume I: alphabetical listing/volume II: appendices, references*, John Wiley & Sons, vol. 41, 2009.
- 79 K. P. Singh and S. Gupta, Nano-QSAR modeling for predicting biological activity of diverse nanomaterials, *RSC Adv.*, 2014, 4(26), 13215–13230.
- 80 F. S. Su and C. S. Lu, Adsorption kinetics, thermodynamics and desorption of natural dissolved organic matter by multiwalled carbon nanotubes, *J. Environ. Sci. Health, Part A: Toxic/Hazard. Subst. Environ. Eng.*, 2007, 42, 1543–1552.
- 81 Z. W. Wang, C. L. Liu, Z. G. Liu, H. Xiang, Z. Li and Q. H. Gong, π - π interaction enhancement on the ultrafast third-order optical nonlinearity of carbon nanotubes/polymer composites, *Chem. Phys. Lett.*, 2005, 407, 35–39.
- 82 F. Tournus, S. Latil, M. I. Heggie and J. C. Charlier, π -stacking interaction between carbon nanotubes and organic molecules, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2005, 72(1–5), 75431.
- 83 S. B. Fagan, A. G. S. Filho, J. O. G. Lima, J. M. Filho, O. P. Ferreira, I. O. Mazali, O. L. Alves and M. S. Dresselhaus, 1, 2-Dichlorobenzene interacting with carbon nanotubes, *Nano Lett.*, 2004, 4, 1285–1288.
- 84 S. Gotovac, Y. Hattori, D. Noguchi, J. Miyamoto, M. Kanamaru, S. Utsumi, H. Kanoh and K. Kanek, Phenanthrene adsorption from solution on single wall carbon nanotubes, *J. Phys. Chem. B*, 2006, 110, 16219–16224.
- 85 J. Zhao and J. Lu, Noncovalent functionalization of carbon nanotubes by aromatic organic molecules, *Appl. Phys. Lett.*, 2003, 82, 3746–3748.
- 86 X. J. Li, W. Chen, Q. W. Zhan, L. M. Dai, L. Sowards, M. Pender and R. R. Naik, Direct measurements of interactions between polypeptides and carbon nanotubes, *J. Phys. Chem. B*, 2006, 110, 12621–12625.
- 87 A. M. Li, Q. X. Zhang, H. S. Wu, Z. C. Zhai, F. Q. Liu, Z. H. Fei, C. Long, Z. L. Zhu and J. L. Chen, A new amine-modified hypercrosslinked polymeric adsorbent for removing phenolics compounds from aqueous solutions, *Adsorpt. Sci. Technol.*, 2004, 22, 807–819.
- 88 W. Chen, L. Duan, L. Wang and D. Zhu, Adsorption of hydroxyl-and amino-substituted aromatics to carbon nanotubes, *Environ. Sci. Technol.*, 2008, 42(18), 6862–6868.
- 89 W. Chen, L. Duan and D. Zhu, Adsorption of polar and nonpolar organic chemicals to carbon nanotubes, *Environ. Sci. Technol.*, 2007, 41(24), 8295–8300.
- 90 L. R. Radovic, C. Moreno-Castilla and J. Rivera-Utrilla, Carbon materials as adsorbents in aqueous solutions, *Chem. Phys. Carbon*, 2001, 227–406.
- 91 C. A. Hunter and J. K. M. Sanders, The nature of π - π interactions, *J. Am. Chem. Soc.*, 1990, 112, 5525–5534.
- 92 J. C. Ma and D. A. Dougherty, The cation- π interaction, *Chem. Rev.*, 1997, 97, 1303–1324.
- 93 C. A. Hunter, K. R. Lawson, J. Perkins and C. J. Urch, Aromatic interactions, *J. Chem. Soc., Perkin Trans. 1*, 2001, 651–669.
- 94 A. H. Sulaymon and K. W. Ahmed, Competitive adsorption of furfural and phenolic compounds onto activated carbon in fixed bed column, *Environ. Sci. Technol.*, 2008, 42, 392–397.



Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table based descriptors

Joyita Roy, Probir Kumar Ojha & Kunal Roy

To cite this article: Joyita Roy, Probir Kumar Ojha & Kunal Roy (2019): Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table based descriptors, *Nanotoxicology*, DOI: [10.1080/17435390.2019.1593543](https://doi.org/10.1080/17435390.2019.1593543)

To link to this article: <https://doi.org/10.1080/17435390.2019.1593543>



Published online: 02 Apr 2019.



Submit your article to this journal [↗](#)



View Crossmark data [↗](#)

Risk assessment of heterogeneous TiO₂-based engineered nanoparticles (NPs): a QSTR approach using simple periodic table based descriptors

Joyita Roy, Probir Kumar Ojha and Kunal Roy 

Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata, India

ABSTRACT

Nowadays, the risk assessment of engineered nanoparticles (NPs) on human health and animals is of great importance. We have used here simple periodic table based descriptors for mixture compounds to predict the cytotoxicity for the heterogeneous NPs. We have developed mono parametric quantitative structure-toxicity relationship (QSTR) models for 34 TiO₂-based NPs modified with (poly) metallic clusters of noble metals (Au, Ag, Pt) to assess the cytotoxicity (-log EC₅₀) towards Chinese Hamster Ovary cell line. After critical statistical analysis of the developed five linear regression (LR) models, we found that the derived models are close to each other in terms of different metric values ($R^2 = 0.922$ – 0.926 ; $Q^2 = 0.907$ – 0.911 ; $R^2_{adj} = 0.918$ – 0.922 ; $Q^2_{F1} = 0.930$ – 0.938 ; $Q^2_{F2} = 0.924$ – 0.932). Thus, we have developed a partial least squares (PLS) model using the five descriptors obtained from the five LR models. The developed PLS model showed good predictivity and robustness in terms of both internal ($R^2 = 0.925$; $Q^2 = 0.911$) and external validation ($Q^2_{F1} = 0.944$; $Q^2_{F2} = 0.938$) parameters. The descriptors, Electrochemical equivalent (E_q), 2nd ionization potential ($2\chi_{pi}$), covalent radius (R_c), amount of Ag (Ag_{amt}) and thermal conductivity (T_c) obtained from the final PLS model well explained the cause of cytotoxicity of the heterogeneous NPs without requiring any computationally expensive descriptors. The insights obtained from the developed models suggested that higher electronegativity, lower oxidation state, and release of metal cation from its oxide increase cytotoxicity through various mechanisms. Thus, these models can be used as efficient tools to assess the toxicity with physiological property of the new heterogeneous NPs in the future.

ARTICLE HISTORY

Received 25 January 2019
Revised 9 February 2019
Accepted 7 March 2019

KEYWORDS


Heterogeneous nanoparticles; QSTR; periodic table descriptors; cytotoxicity

1. Introduction

Nanotechnology has taken the frontline in the modern world of science (Islam and Miyazaki 2010). Nanoscale materials are substances comprising one or more features less than 100 nm in at least one dimension. In theory, nanoparticles can be engineered from any substance like semiconductors nanocrystals, organic dendrimers, and carbon fullerenes, and they possess properties like electrical, thermal, mechanical which are highly desirable in commercial, medical and environmental sectors (National Research Council 2002). There are two major classes of nanoparticles: (a) carbon nanoparticles (b) metal oxide nanoparticles (Gajewicz et al. 2012). Metal oxide nanoparticles (NPs) have unique property due to their size and high density of edge or corner sites. There is an increase in the use of

nanoparticles in different areas like space technology, pharmacy, environmental engineering, cosmetic, stain-resistant clothing, environmental monitoring, and so on (Artiles et al. 2011; Puzyn et al. 2010). By the end of 2019, their worldwide market is estimated to be \$79.8 billion (Highsmith 2014). It has been noted that metal oxides induce toxicity to some organisms (Dreher 2004), and it is believed that an exponential growth in the use of nanoparticles may endanger human health through the potential induction of cytogenetic, mutagenic, or neurotoxic health effects (Cattaneo et al. 2010; Hansen et al. 2008). The assessment and characterization of risk posed by metal nanoparticles to human beings and animals are complex, as they have a wide range of shapes, sizes, surface modifications, and different chemical compositions, and

CONTACT Kunal Roy  kunalroy_in@yahoo.com, kunal.roy@jadavpuruniversity.in; Probir Kumar Ojha  probirojha@yahoo.co.in  Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

 Supplemental data for this article can be accessed [here](#).

© 2019 Informa UK Limited, trading as Taylor & Francis Group

all of these may affect toxicity towards the biological cells. The heterogeneous NPs ($\text{Me}_{\text{mix}}@\text{MeO}_x$, where Me_{mix} is a bimetallic cluster and MeO_x is a metal oxide nanoparticle) could serve as photocatalysts for degrading a variety of inorganic and organic compounds (Serpone and Emeline 2012) and can be used in visible light induced process to remove the hazardous pollutants from both aqueous and gaseous medium (Mikolajczyk et al. 2016). Heterogeneous NPs are of great interest to the industries but they too possess adverse impact on the environment and human beings (Handy et al. 2008a; Handy et al. 2008b). Hence, a great deal of work is going on to develop a defined method for characterization, engineering control, transport, fate as well as an assessment of their life cycle and exposure of the metal oxide NPs (Morris et al. 2011). The need to develop a novel, cost-effective and fast method to minimize the toxic behavior, properties and environmental impact of metal oxide nanoparticles is stressed in the REACH legislature (European 2006), the European Chemical Agency (ECHA 2012), United States-Canadian Regulatory Cooperation Council (RCC-NI 2011), and also OECD (Organization for Economic Co-operation Development) (OECD 2014) in order to cope up with the situation of metal oxide toxicity. The proposed alternative procedure (method) should not only be able to perform risk assessment but also to reduce the extensive animal testing and provide detailed information about toxicity mechanism at the molecular level. As the conventional process or method of risk assessment is expensive and time-consuming, a computational-based assessment method should be alternatively used which not only reduces the number of experiments but also the cost of consumable reagents.

Different computational methods have been developed over the past several years, but quantitative structure-activity/property/toxicity relationship (QSAR/QSPR/QSTR) models originally introduced by Corwin Hansch for the first time in 1962 is the promising one. The fundamental assumption of QSAR is that the molecular structures are responsible for the chemical, physical and biological properties of the studied samples (Tropsha 2003). Each compound can be described in terms of descriptors; for this purpose, different features of chemical structures are represented within a mathematical

context (matrix or coordinates) and then transformed into numbers or descriptors (Puzyn et al. 2010; Todeschini and Consonni 2009). Nano-QSAR is based on the mathematical dependency between the variance in the molecular property (encoded by nano descriptors) and biological activity (Puzyn et al. 2011; Zhu et al. 2009). This means that one can easily predict the missing data from the theoretically generated descriptors, if one has access to the toxicological data and calculated descriptors for a group of homogeneous or alike nanoparticles. QSAR models work on the three R's (3Rs) principle, i.e. replacement, refinement and reduction in animal research (Benigni and Giuliani 2003). In the past few years, the development of nano-QSAR models has been impacted by the introduction of the perturbation theory. Its main aim was to predict simultaneously ecotoxicity and cytotoxicity of NPs against various assay organisms (plants, crustaceans, algae, bacteria, and cells lines, fungi, nematodes, amphibians, fishes etc.) by taking into consideration of multiple measures of ecotoxicity (EC_{50} , IC_{50} , LC_{50} , TC_{50} , CC_{50}) at under various experimental conditions, including varying exposure time, multiple biological targets, compositions, different sizes, diverse measures of toxicities and also conditions to measures those sizes, shapes, times during which the biological targets were exposed to NPs and coating agents (Concu et al. 2017; Kleandrova et al. 2014a; Kleandrova et al. 2014b; Luan et al. 2014; Speck-Planche et al. 2015). QSAR has been proven to be an efficient tool for predicting the adverse effects of chemical entities in terms of chemical screening, risk assessment, and priority setting (Roy and Das 2013).

In the present work, we have explored the development of predictive models for cytotoxicity ($-\log \text{EC}_{50}$) of metal oxide nanoparticles to the Chinese hamster ovary cells using a data set of 34 modified TiO_2 NPs. This data set was also previously modeled by Mikolajczyk et.al. (2018) using quantum chemical descriptors. Calculation of quantum descriptors is computationally demanding. Hence, we have attempted to develop nano-QSTR models here using periodic table based descriptors and extensively validated the models with different strategies. The periodic table based descriptors can be calculated involving less computational resources, and they can provide good interpretability.

1.1. Integral additive descriptors

According to the results from various works of literature (Hewlett and Wilkinson 1967), additive schemes are appropriate for the preliminary modeling of multiple mixture chemical toxicity. The modes of action (generally four major types) for some mixture of conventional organic compounds were defined previously as, (a) simple additive, i.e. the combined toxic response is equal to the total of the single chemical toxicity, (b) greater than synergism or additive that means that the joint effect is more than the sum of the toxicity of individual chemicals, (c) the total toxicity is less than the overall individual chemical toxicity, i.e. less than additive or partial additive, (d) independent or no interaction which means the combined toxic effect is equal to that caused by the component with highest toxicity (Hewlett and Wilkinson 1967; Plackett and Hewlett 1967). The joint effect of the non-reactive chemicals might be significantly different from the scheme of simple addition as studied from the previous literature (Hewlett and Wilkinson 1967). Hence, we have decided to develop nano-QSTR models using additive descriptors which were first explored by Mikolajczyk et al (2018) for the heterogeneous TiO₂ based modified nanoparticles. We have assumed here that each individual component contributes additively to the toxicity, and its contribution is proportional to the individual component mole fraction in the mixture (Hewlett and Wilkinson 1967; Plackett and Hewlett 1967). Summation of the concentration of the individual components in the mixture was carried out after multiplying each with a scaling factor that indicates the contribution of property of the individual components (C_i), and hence the sum of the concentration (a) is the property of mixture based NPs (C_{mix}).

$$C_{mix} = \sum_{i=1}^n aC_i \quad (1)$$

Here, we have applied the concept of integral additive descriptors which was previously proved efficient for modeling the heterogeneous organic mixture systems (Altenburger et al. 2004; Melagraki and Afantitis 2014; Roy and Ambure 2016; Tropsha 2010; Wang et al. 2014; Xu and Nirmalakhandan 1998; Zhang et al. 2007) and used them to describe the influence of the amount of the noble metal in

the structure of the TiO₂ NPs. The individual component can be expressed as a set of 2 D and 3 D descriptors in the framework of integral additive scheme, and the descriptor is expressed as mole weighted average and mole fraction of each component as follows:

$$D_{mix} = R_1D_1 + R_nD_n \quad (2)$$

where D_{mix} corresponds to the mixture descriptor, R_1 and R_n represent the mole fraction of the individual component in the mixture, and D_1 and D_n stands for descriptors of each component in the mixture.

2. Methods and materials

2.1. Data set

In the present work, we have developed nano-QSTR (Nano-Quantitative Structure-Toxicity Relationship) models for 34 TiO₂ (Mikolajczyk et al. 2018) NPs (nanoparticles) modified with varying amount and types of mixture of noble metals like Ag, Au and Pt (expressed in mole %). The cytotoxicity data towards the Chinese hamster ovary cell line (CHO-K1, ATCC[®] CCL-61TM) was expressed in $-\log EC_{50}$ (negative logarithm of EC_{50}) for the development of Nano-QSTR models. For the purpose of modeling, all the nanoparticles were utilized and no single NPs were omitted. All the NPs used in the Nano-QSTR modeling were obtained from microemulsion method (Mikolajczyk et al. 2018). The details of the dataset are given in the Supplementary section (Table S1).

2.2. Descriptor calculation

The Nano-QSTR models were developed from the fundamental information of the noble metals obtained from the periodic table to relate it to the toxicity towards hamster ovary cell line and to investigate the modified TiO₂ based NPs in order to identify the key structural features responsible for the toxicity. For this purpose, we have taken the information about different properties of various metals used to modify the surface of the NPs directly from the periodic table. These properties were converted to integral additive descriptors which were used for model development purpose as described later in this section. Note that, we have

generated 30 descriptors directly or indirectly from the periodic table derived information, and three descriptors were calculated based on the number of different metals used to modify the surface of TiO₂ based NPs. The used periodic table based descriptors are given in the Supplementary section (Table S2). The periodic table descriptors can be adopted for calculation of integral additive descriptors of modified TiO₂ based NPs. To understand the structural changes in the heterogeneous NPs after modification of TiO₂ NPs using single metal clusters or with varying amount and types of noble metals with different concentration, the modified form of Equation (2) has been used here for model development purpose. The calculation of the proposed mixture descriptors used in this work can be represented by the following equation,

$$D_{\text{mix}} = \% \text{ mol}_{\text{Me}(1)} \times P_1 + \dots + \% \text{ mol}_{\text{Me}(n)} \times P_n \quad (3)$$

where, D_{mix} means mixture descriptor, $\% \text{ mol}_{\text{Me}1}$ means concentration of each metal/component in the mixture – contribution by weight of metal in the NP sample of the mixture and P_n means the periodic table descriptor of an individual metal. This method is used in order to treat each individual metal in the cluster as a mixture system, and each metal is described as a set of descriptors (calculated from the periodic table). The descriptor selected for the metal (used for coating of TiO₂ based NPs) from the periodic table is multiplied with the amount of the metal present in the mixture (TiO₂ based modified NPs), and the resulting sum defines a new set of descriptors for the complex mixture system. Periodic table based descriptors are more advantageous than others as they are easy to obtain without any significant calculation unlike quantum chemical descriptors (De et al. 2018; Kar et al. 2014; Kar et al. 2016). The list of periodic table descriptors is given in the Supplementary section (Table S3).

2.3. Data set division

Data set division is a very crucial step for development of Nano-QSTR models. The data sets in the present work were divided into training (75% of the total dataset compounds) and test (25% of the total dataset compounds) sets based on “Modified k-Medoid” clustering technique. This method

categorizes compounds into clusters so that the compounds in the same cluster are similar and compounds from different clusters are dissimilar. This method tends to select the “k” most centered compounds or objects as the initial Medoids. Here, three clusters were obtained for 34 NPs. The purpose of the training set is to develop the models, and the test set is utilized for validating the models in terms of significance and robustness.

2.4. Model development

After performing the data set division, we have developed mono parametric Nano-QSTR models employing the Best Subset Selection v2.1 software (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab) using mixed descriptors (33 descriptors) and selected five mono parametric LR models based on the MAE based criteria. After that, we have performed PLS regression of all the descriptors obtained from the previously developed five linear regression (LR) models with one latent variable.

2.4.1. Model selection

The selection of the best descriptors for mono-parametric models is performed using software developed in our laboratory (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab). The evaluation of the additive descriptors of mixture system was done from the evaluation of different statistical parameters (R^2 , Q^2 , Q^2_{F1} , Q^2_{F2}) (Chirico and Gramatica 2011). In the present work, we performed the Best Subset Selection v2.1 with the 34 periodic table descriptors and selected five LR models based on the MAE based criteria (Roy et al. 2016).

2.4.2. Chemometric tools

The mono-parametric Nano-QSTR models were developed using the Best subset selection software v2.1 available from our laboratory (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab). Partial least squares (PLS) regression (Chin 1998) was used for the development of the final model using all the individual descriptors obtained from the LR models. In case of PLS regression, overfitting is avoided by checking the number of PLS components; if a new PLS component is insignificant then the PLS run is stopped. Clustering of the data set was performed by

“Modified K-Medoid” tool version 1.3 (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab) for splitting the data sets into a training set and a test set.

2.5. Validation metrics

To determine the predictability and reliability of the models, internal (R^2 , R^2_{ar} , Q^2_{LOO} , variance ratio (F), and standard error of estimate(s)) and external (R^2_{pred} , Q^2_{F2} and concordance correlation coefficient (CCC)) validation metrics were computed. Additionally, r_m^2 parameters like $r_m^2_{(LOO)}$ and $\Delta r_m^2_{(LOO)}$ for internal validation and $r_m^2_{(test)}$ and $\Delta r_m^2_{(test)}$ for external validation (Ojha et al. 2011) were also calculated. We have also judged the models based on MAE (mean absolute error) criteria, as the R^2_{pred} and Q^2_{F2} (Q^2_{ext} based metrics) are not always able to provide a true indication of the prediction potential, as they are influenced by the data range and distribution of the response value in the training set and the test set (Roy et al. 2016). The purpose of the predictive performance of QSAR models is to detect the prediction errors of the test set (external set).

2.6. Applicability domain (AD)

Technically, AD (applicability domain) represents the chemical space suggested by the structural

information of the chemicals used in model development (the training set compounds) in a QSTR analysis. Here, we have used DModXP approach (Umetrics 2013) for the PLS model and Standardization approach (Roy et al. 2015) for LR models to determine whether the test compounds under consideration is within or outside the AD.

The methods involved for the development of final model are schematically represented in Figure 1.

3. Results and discussion

Based on the cytotoxicity data of 34 TiO_2 modified NPs towards Chinese hamster ovary cells and easily calculated 34 periodic table based descriptors, we have developed five simple but statistically significant LR based Nano-QSTR models. We have checked both the internal ($R^2 = 0.922$ – 0.926 ; $Q^2 = 0.907$ – 0.911 ; $R^2_{adj} = 0.918$ – 0.922) and external ($Q^2_{F1} = 0.930$ – 0.938 ; $Q^2_{F2} = 0.924$ – 0.932) validation parameters for all the individual models which showed good *in silico* predictivity of the models as depicted in Table 1. For the external validation, Q^2_{F1} or R^2_{pred} and Q^2_{F2} metrics were used, and their values are much higher than the threshold value, i.e. 0.5. The MAE based criteria were also checked for the models, and each individual model passed the MAE

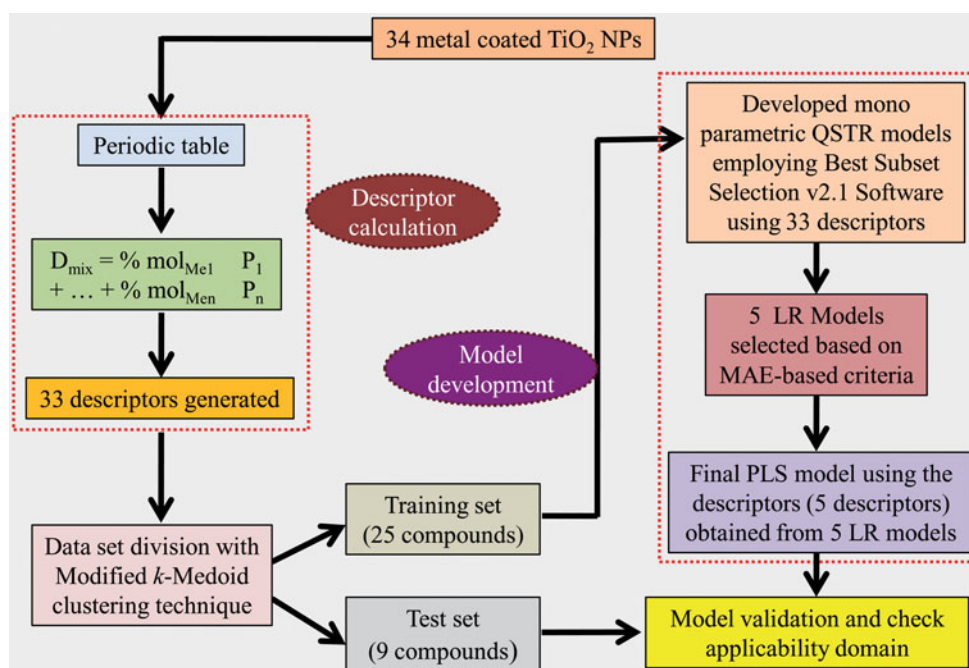


Figure 1. Schematic representation of the steps involved for the development of QSTR models.

Table 1. Statistical quality and validation parameters obtained from the developed LR and PLS models.

Types of models	Descriptors	Training set statistics					Test set statistics						
		Model R ²	Model Q ² _(LOO)	MAE _(95%)	r ² _{m(LOO)}	Δr ² _{m(LOO)}	R ² _{pred} or Q ² _{F1}	Q ² _{F2}	CCC	r ² _{m(test)}	Δr ² _{m(test)}	MAE _(95%)	MAE
LR	M1 Electrochemical Equivalent (g/amp-h)	0.926	0.911	Moderate	0.880	0.045	0.937	0.931	0.959	0.828	0.049	0.066	Good
	M2 2nd ionization (kJ/mol)	0.923	0.909	Moderate	0.877	0.047	0.938	0.932	0.960	0.843	0.042	0.057	Good
	M3 covalent radius (pm) https://doi.org/10.1073/pnas.90.17.7915	0.923	0.909	Moderate	0.877	0.047	0.938	0.932	0.960	0.842	0.042	0.056	Good
	M4 Amount of Ag	0.922	0.907	Moderate	0.875	0.048	0.930	0.924	0.956	0.806	0.059	0.073	Good
	M5 Thermal conductivity(W/(m. k))	0.922	0.907	Moderate	0.874	0.048	0.934	0.928	0.958	0.829	0.052	0.077	Good
PLS	P1 All 5 descriptors with one latent variable	0.925	0.911	Moderate	0.883	0.0483	0.944	0.938	0.969	0.922	0.031	0.068	Good

based criteria. We have calculated r_m^2 parameters like $\overline{r_{m(loo)}^2}$ and $\Delta r_{m(loo)}^2$ and $Q_{(LOO)}^2$ for the internal set and $r_{m(test)}^2$ and $\Delta r_{m(test)}^2$ for the external set, and the resultant values passed the critical values proving statistical reliability of the models. The descriptors obtained from the five LR models are discussed elaborately below. In the presented equations, n_{training} means number of compounds in the training set/internal set used to develop the models, and n_{test} means number of compounds in the test set/external set used to judge the quality of the models. The test compounds showed good predictivity based on the models. The scatter plot of observed vs predicted cytotoxicity for all the LR and PLS models are shown in Figure 2.

Model 1

The mono parametric equation is as follows:

$$-\log EC_{50} = 4.673(\pm 0.034) + 0.458(\pm 0.003)Eq$$

$$n_{\text{training}} = 25, R^2 = 0.926, R_{\text{adj}}^2 = 0.922, S = 0.127, \\ F = 286.07,$$

$$\text{PRESS} = 0.440, Q^2 = 0.911, \overline{r_{m(LOO)}^2} = 0.880, \Delta r_{m(LOO)}^2 \\ = 0.045, \text{MAE based criteria} = \text{Moderate},$$

$$n_{\text{test}} = 9, Q_{F1}^2 = 0.937, Q_{F2}^2 = 0.931, \overline{r_{m(test)}^2} \\ = 0.828, \Delta r_{m(test)}^2 = 0.049, \text{MAE based criteria} \\ = \text{Good}$$

Electrochemical equivalent (Eq)

The descriptor Electrochemical Equivalent (g/amp-hr) (Eq) of a chemical element indicates the mass of that element (in grams) transported by 1 coulomb of electric charge. Electrochemical Equivalent can be calculated as follows,

Electrochemical Equivalent = Gram molecular mass of the substance/number of electrons (involved in reaction)

This indicates that the atoms containing lower number of valence shell electrons will have higher descriptor values, as it is inversely proportional to

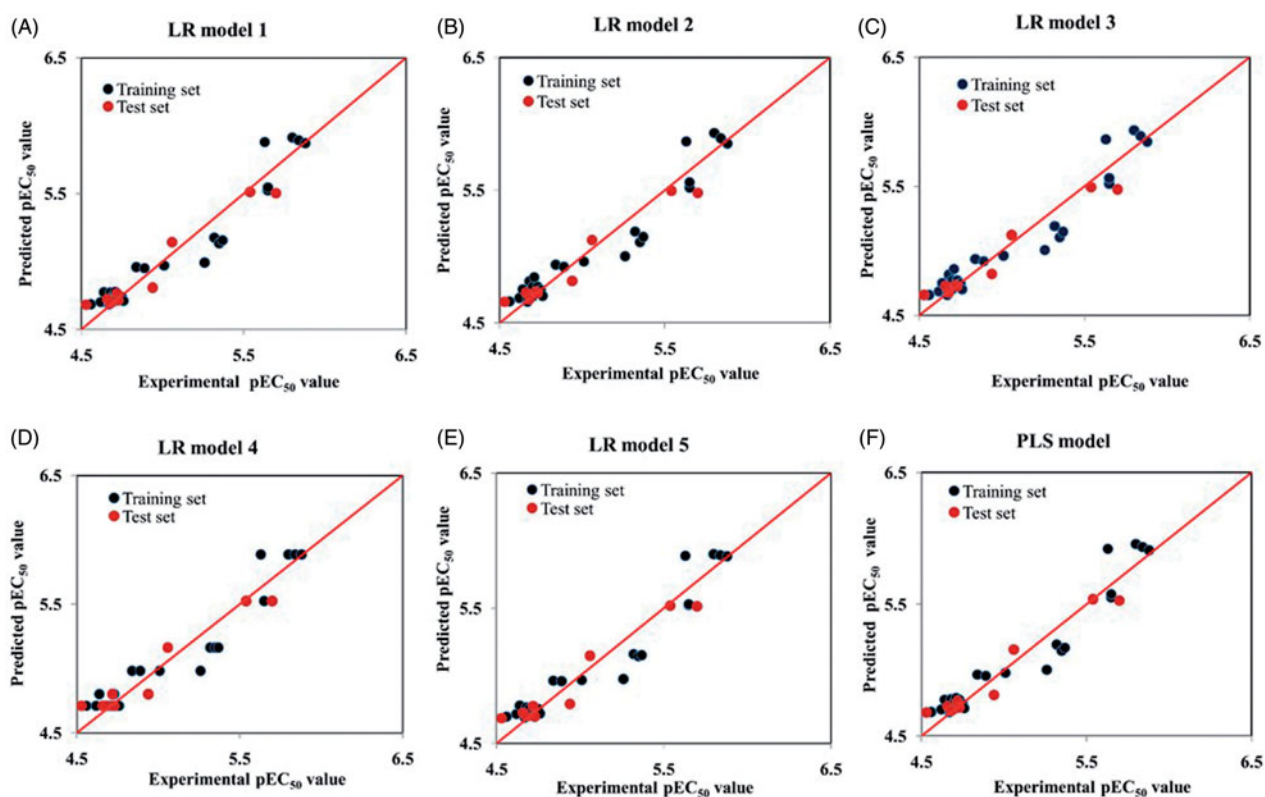


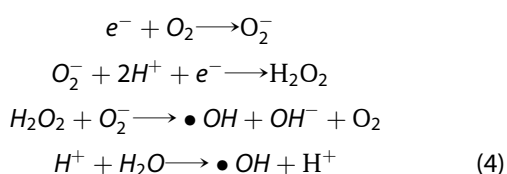
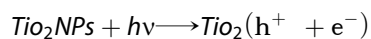
Figure 2. The scatter plot of the observed and the predicted cytotoxicity ($-\log EC_{50}$) values of the LR and PLS models. Figure (A-E) represents LR models and Figure F represents PLS model.

the number of electrons in the principle valence shell taking part in the chemical reaction. The descriptor also determines the kinetics of corrosion rate and estimates the oxidizing power of metal in specific environment. The oxidizing power can be determined through the oxidation potential of a metal, which gives the measure of the likelihood of a metal to move from lower oxidation state to higher oxidation state. The transition metals can exist in different oxidation state as they have partially filled $-d$ and $-f$ orbital shells. The elements with low number of valence electrons will have less oxidation state and metals with less oxidation state are more harmful than the elements with higher oxidation or stable oxidation state (Walker et al. 2003). The positive regression coefficient of this descriptor suggests that the toxicity towards hamster ovary cell will increase with an increase in the numerical value of this descriptor as shown in case of nanoparticles 6.5Ag_{0.5}Pt and 6.5Ag_{0.25}Pt (the Electrochemical Equivalent values are 27.06975 and 26.614825 (g/amp-hr) and their corresponding toxicity values are 5.8 and 5.84 respectively) and vice

versa as shown in nanoparticles 0.05Au_{0.05}Pt and 0.1Au (the Electrochemical Equivalent values are 0.213465, 0.24496 (g/amp-hr) and their respective toxicity values are 4.67 and 4.56).

Mechanism of toxicity

The toxicity of a metal ion depends on its electrochemical features, solubility, and stability. Chelating ability of the metal ion with the particular ligands of biological macromolecules also affect the toxicity to the biological cells. Toxicity of the metal depends both quantitatively and qualitatively on the oxidation state of the metals. The lower oxidation state metals are more toxic than their higher oxidation state due to its tendency to get oxidized to form stable oxides, i.e. higher valence state hence disrupting cellular processes (Walker et al. 2003). Electron detachment from metal NPs initiates lipid peroxidation by reactive oxygen species (ROS) such as superoxide ($O_2^{\bullet-}$) and hydroxyl radicals ($\bullet OH$) (Lovrić et al. 2005; Neal 2008)). Using TiO_2 NPs as example, ROS is produced as per the following scheme in presence of light radiation:



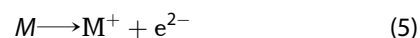
The cellular damage as per true toxicity mechanism may involve release of metal ions. The extent of ROS production increases by direct contact of the nanoparticles (NPs) with the cell (Kang et al. 2007). NPs can increase the oxidative stress as per the given mechanism by generating reactive oxygen species (ROS) which reduces the antioxidants (Lin et al. 2006b; Lin et al. 2006a) that eventually leads to cell injury and death of the cell. The production of the high energy species may attack lipids, proteins, nucleic acid or other biological macromolecules thus causing damage to the cells. They may hamper the mitochondrial structure and depolarize the membrane, even may cause impairment of the electron transport chain and activation of the NADPH system (Xia et al. 2006). Damage to the DNA may lead to cell cycle arrest and apoptosis.

Model 2

$$\begin{aligned} -\log EC_{50} &= 4.643(\pm 0.036) + 0.0001(\pm 0.00001)2\chi_{pi} \\ n_{\text{training}} &= 25, R^2 = 0.923, R_{\text{adj}}^2 = 0.920, S = 0.129, \\ &F = 277.15, \\ \text{PRESS} &= 0.452, Q^2 = 0.909, \overline{r_{m(\text{LOO})}^2} = 0.877, \\ \Delta r_{m(\text{LOO})}^2 &= 0.047, \text{MAE based criteria} = \text{Moderate}, \\ n_{\text{test}} &= 9, Q_{F1}^2 = 0.938, Q_{F2}^2 = 0.932, \overline{r_{m(\text{test})}^2} \\ &= 0.843, \Delta r_{m(\text{test})}^2 = 0.042, \text{MAE based criteria} = \text{Good} \end{aligned}$$

2nd ionization potential (${}^2\chi_{pi}$)

The next significant descriptor, 2nd ionization potential (${}^2\chi_{pi}$), also contributes to the cytotoxicity of the hamster ovary cell. This descriptor defines the energy needed to remove a second electron from each ion in one mole of gaseous 1^+ ion to give gaseous 2^+ ions.



Here, M is the atom, M^+ and M^{2+} are the metal ions and e^- is the electron, i.e. ionization energy.

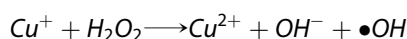
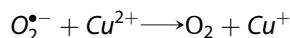
Ionization potential is the difference of energy between the ground state and state of ionization, and this amount of energy is required to completely remove the loosely attached electrons. The 2nd ionization potential is greater than 1st ionization potential and depends upon the size, charge and the type of electrons removed from outer shell of the atom. Ionization potential also determines the electronegativity and electron affinity of an atom. The low ionization energy of an atom (the energy required to remove the outer shell electron) indicates that the atom can easily lose its outer shell electron and has fewer tendencies to gain electrons. Thus, it clearly indicates that the atoms with high ionization potential will have high electronegativity. The electronegativity is responsible for the catalytic property of the cationic form of the metal and therefore increases the cytotoxicity. The positive regression coefficient of this descriptor indicated that an atom with higher 2nd ionization potential increases the cytotoxicity of the hamster ovary cell and vice versa. As for example, the nanoparticles 6.5Ag_0.5Pt and 6.5Ag are highly toxic (toxicity values are 5.8 and 5.88 respectively) towards the cytotoxicity to hamster ovary cell due to their higher range of 2nd ionization potential (14350.5 and 13455 respectively), whereas in case of nanoparticles 0.25Pt and 0.1Au, the cytotoxicity (4.56 and 4.67 respectively) decreases with its 2nd ionization potential (447.75 and 198 kJ/mol respectively).

Mechanism of toxicity

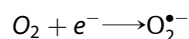
Electronegativity depends on the atomic radius and on the formal charge of the cationic metal. Metal nanoparticles containing higher electronegativity have a tendency to gain electrons from the bonding pair of the electrons. Therefore, an increase in electronegativity suggests an increase in the catalytic properties of the cationic metal, and thus it increases the toxicity of the metal nanoparticles as described by the Haber–Weiss–Fenton cycle (Koppenol 2001). Electronegativity reduces with the number of valence electrons.

Electronegativity of the metal separates the metal cation from the metal oxide NPs during the toxic effect. Oxidative stress caused here due to generation of intracellular ROS levels causes oxidative damage to the cells leading to apoptosis (Ozben 2007). The number of ROS as OH radicals (Cho et al. 2004), superoxide ions (Sawai et al. 2000), hydrogen peroxide (H_2O_2) is found to be responsible for the generation of oxidative stress in the cell.

The Haber–Weiss–Fenton cycle is explained using copper metal as an example:

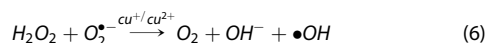


Usually, $\bullet OH$ radicals are produced in all aerobic organisms in the form of byproducts of cellular respiration as they use oxygen (molecular) to obtain energy. The problem arises when there is an imbalance between the oxidative and reductive products due to increase ROS production.



Superoxide anion radicals are products of one-electron reduction of the O_2 atom.

The electrons required for this reaction is utilized from electron transfer chain at the time of cellular respiration (Ames et al. 1993).



Thus, in the presence of the metal cation, the hydroxyl radical ($\bullet OH$) is formed more readily than the normal. The high concentration of the hydroxyl radical becomes elevated than natural scavengers in the cell, which causes an imbalance in the antioxidants in the cell, ultimately leading to oxidative stress and cell death.

Model 3

$$\begin{aligned} -\log EC_{50} &= 4.641(\pm 0.036) + 0.0013(\pm 0.0001)R_c \\ n_{\text{training}} &= 25, R^2 = 0.923, R_{\text{adj}}^2 = 0.920, S = 0.129, \\ F &= 276.74, \\ \text{PRESS} &= 0.452, Q^2 = 0.909, \overline{r_{m(\text{LOO})}^2} = 0.877, \\ \Delta r_{m(\text{LOO})}^2 &= 0.047, \text{MAE based criteria} = \text{Moderate}, \\ n_{\text{test}} &= 9, Q_{F1}^2 = 0.938, Q_{F2}^2 = 0.932, \overline{r_{m(\text{test})}^2} \end{aligned}$$

$$= 0.842, \Delta r_{m(\text{test})}^2 = 0.042, \text{MAE based criteria} = \text{Good}$$

Covalent radius (R_c)

The covalent radius (R_c) descriptor is a measure of the size of an atom that forms a part of one covalent bond, and it is the third common measure of the size of the atom. It is primarily calculated from the nuclear charge, i.e. atomic number and electronic configuration of the atom. The positive regression coefficient of the descriptor suggests that the numerical value of this descriptor is directly correlated with the cytotoxicity as shown in nanoparticles 6.5Ag_0.1Pt and 6.5Ag, where the cytotoxicity (5.63 and 5.88 respectively) increases with covalent radius of nanoparticles (956.1(pm) and 942.5(pm) respectively). On the other hand, when the covalent radius of the nanoparticles decreases as in case of nanoparticles 0.25Pt (34 pm) and 0.1Au_0.25Pt (47.6 pm), the respective toxicity 4.67 and 4.7 also decreases. If the size of atomic radius increases, the number of shells also increases, shielding the outer electrons from the electrostatic pull of the nucleus. Again, the outer valence shells can easily lose electrons to form cation radical that may cause further modification to DNA bases and enhance lipid peroxidation ultimately causing cytotoxicity.

Mechanism of toxicity

The presence of metal cations instigates the formation of sufficient amount of radicals ($\bullet OH$) than the metals do naturally. The elevated level of the reactive hydroxyl radical makes it impossible for the natural scavengers to keep the normal physiological balance in the cell. The metal cation increases the production of the free radicals both in cell and mitochondria which attack the DNA and mtDNA (mitochondrial DNA) respectively and causes fragmentation of the DNA. When metal radical attacks the protein, it causes blockade of the protein synthesis by oxidizing them leading to autocatalytic lipid peroxidation. This mechanism decreases the mitochondrial membrane potential, which leads to the loss of the mitochondrial membrane fluidity, and thus the content of the matrix is spilled out into the inner membrane

Model 4

$$\begin{aligned}
 -\log EC_{50} &= 4.709(\pm 0.338) + 0.181(\pm 0.011)\text{Amount of Ag} \\
 n_{\text{training}} &= 25, R^2 = 0.922, R_{\text{adj}}^2 = 0.919, S = 0.130, \\
 F &= 272.26, \\
 \text{PRESS} &= 0.459, Q^2 = 0.907, \overline{r_{\text{m(LOO)}}^2} = 0.875, \Delta r_{\text{m(LOO)}}^2 \\
 &= 0.048, \text{MAE based criteria} = \text{Moderate}, \\
 n_{\text{test}} &= 9, Q_{F1}^2 = 0.930, Q_{F2}^2 = 0.924, \overline{r_{\text{m(test)}}^2} \\
 &= 0.806, \Delta r_{\text{m(test)}}^2 = 0.059, \text{MAE based criteria} = \text{Good}
 \end{aligned}$$

Amount of Ag (Ag_{amt})

The descriptor, amount of Ag, determines the measurement of silver metal concentration. The positive regression coefficient indicates that with an increase in the amount of the silver metal as shown in case of nanoparticles 6.5Ag_0.5Pt and 4.5Ag_0.5Pt (6.5 and 4.5 mol %_respectively), the cytotoxicity (5.8 and 5.65 respectively) of the metal towards the hamster ovary cell also increases. On the other hand, when the amount of silver is reduced as shown in case of nanoparticles 0.5Ag_0.1Pt and 1.5Ag (0.5 and 1.5 mol % respectively), the corresponding toxicity value (4.64 and 4.89 respectively) also decreases, which clearly depicts that silver metal has a dominant role for cytotoxicity towards the hamster ovary cell. Silver metal has an antimicrobial effect and induces toxicity in many types of species (Bilberg et al. 2011) and chronic exposure of silver metal may cause argyria or argyrosis in humans as suggested by some authors (Drake and Hazelwood 2005). Silver metal has a better water solubility than the other metals and thus its concentration is higher in the solution compared to other investigated metals of equal molar mass. Due to the higher concentration of silver metal available in solution, they are more toxic than the other metals.

Mechanism of toxicity

Different hypotheses have been formulated for the mechanism of silver ion toxicity to the cell. Among the various hypotheses, the silver ion release from the metal oxide (Kittler et al. 2010) and generation of the reactive oxygen species (ROS) (Cheng et al. 2013; Foldbjerg et al. 2011) is suggested to be most

likely. Silver NPs are believed to produce toxicity through so-called Trojan-horse mechanism (Lubick 2008). In this mechanism, Ag is released intracellularly after being taken up by the cell and subsequently causes death of the cell. Ag NPs accumulate into the cell and produce ROS directly or may indirectly increase ROS production by reducing antioxidants production. Thus, it decreases the viability of the cells and also induces damage of DNA and chromosomes (Foldbjerg et al. 2011; Jiang et al. 2013; Kim et al. 2010) ultimately leading to apoptosis of the cell. Smaller Ag NPs are more toxic than the larger ones because of their high surface to volume ratio, which further facilitates the release of the Ag ions in the cell.

Model 5

$$\begin{aligned}
 -\log EC_{50} &= 4.682(\pm 0.035) + 0.0004(\pm 0.000026)Tc \\
 n_{\text{training}} &= 25, R^2 = 0.922, R_{\text{adj}}^2 = 0.918, S = 0.130, \\
 F &= 270.44, \\
 \text{PRESS} &= 0.463, Q^2 = 0.907, \overline{r_{\text{m(LOO)}}^2} = 0.874, \\
 \Delta r_{\text{m(LOO)}}^2 &= 0.048, \text{MAE based criteria} = \text{Moderate}, \\
 n_{\text{test}} &= 9, Q_{F1}^2 = 0.934, Q_{F2}^2 = 0.932, \\
 \overline{r_{\text{m(test)}}^2} &= 0.829, \Delta r_{\text{m(test)}}^2 = 0.052, \\
 \text{MAE based criteria} &= \text{Good}
 \end{aligned}$$

Thermal conductivity (T_c)

Thermal conductivity (T_c) is a property of metals. It determines the rate at which heat passes through a particular material; it is expressed as the amount of heat that passes through unit area per unit time and possesses temperature gradient per degree per unit distance. According to the band theory, the atoms of metal crystals are very close to each other causing the orbitals to overlap each other, suggesting that there is presence of large number of electrons in a small piece of metal and due to their closeness they are referred to as bands. The filled bands are known as valence bands and partially filled bands with delocalized electrons are called conduction bands. Since in metals the closeness is very small, therefore it becomes easy for the electrons to move from valence band to conduction band. The ability of metals to conduct electricity depends on the proximity of the valence and

conduction bands. Band theory also explains the possibility of the movement of delocalized electrons which is due to the overlapping of the molecular orbitals. The positive regression coefficient of this descriptor indicates that the cytotoxicity of the nanoparticles increases with its thermal conductivity and vice versa. It has been observed in case of nanoparticles 6.5Ag_0.25Pt and 6.5Ag that the cytotoxicity of these nanoparticles (5.84 and 5.88 respectively) increases as the thermal conductivity also increases (2806.4 (W/(m•K) and 2788.5 (W/(m•K) respectively), whereas the reverse occurs in the nanoparticles 0.25Pt and 0.1Au where the cytotoxicity (4.67 and 4.56 respectively) decreases with thermal conductivity (17.9 and 31.8 respectively). Thus, increase in the thermal conductivity means that there is a decrease in the band gap, which makes it easier for the movement of electrons to the conduction band, and hence overlapping of the band gaps causes oxidative stress and acute pulmonary inflammation compared to the material whose band gaps does not overlap (Zhang et al. 2012).

Mechanism of toxicity

ROS production and oxidative stress occur due to the band gap of the nanoparticle energy band. The intracellular redox processes occurring in the biological media initiates electron transfer process from the valence band to conduction band. Burello and Worth et.al (2012) stated that redox potential (E_0) of the naturally occurring reaction in the cell in the context with the values of conduction band energy (E_c), and valence band energy (E_v) may be the main reason for the toxicity of the nanoparticle oxides. The overlapping of the E_c and E_v band causes oxidative stress which leads to the imbalance between the production of free radicals and the ability of the body to detoxify or counteract their harmful effects through neutralization by antioxidants. The toxicity arises due to detachment of the electron from the modified metal oxide NPs, i.e. reductive potential. Electron release in the cell interacts with various molecules to produce a free radical ($\bullet OH$) that attacks the DNA double strand and blocks the replication or otherwise block the protein and oxidize them which impairs their function.

When there is a sufficient DNA damage, then the cell undergoes apoptosis.

The PLS model

After critical analysis of the statistical results (both internal and external validation parameters) obtained from the five LR models, we found that all five descriptors were significant in modeling toxicity in the Chinese hamster ovary cell. Therefore, we performed PLS regression with the same data set division using the five descriptors obtained from the LR models. The final PLS equation was developed using only one latent variable.

$$-\log EC_{50} = 4.669 + 0.00918E_q + 0.00002\chi_{pi} + 0.00026R_c + 0.03627 Ag_{amt} + 0.00009 T_c$$

$$N_{train} = 25, R^2 = 0.925, R_{adj}^2 = 0.922, S = 0.127,$$

$$F = 284.75, PRESS = 0.0381, Q_{(LOO)}^2 = 0.911,$$

$$LV = 1, \overline{r_{m(loo)}^2} = 0.883, \Delta r_{m(loo)}^2 = 0.048$$

$$N_{test} = 9, Q_{F1}^2 = 0.944, Q_{F2}^2 = 0.938, \overline{r_{m(test)}^2} = 0.922,$$

$$\Delta r_{m(test)}^2 = 0.031, MAE \text{ based criteria} = \text{Good}.$$

Using the variable importance plot (VIP) (Figure S1), the significance level of the descriptors was found to be in the following order: E_q , $2\chi_{pi}$, R_c , Ag_{amt} , and T_c . Both external and internal validation results showed good predictivity pattern. The Q_{F1}^2 (0.944) metric proves high predictability of the developed model. The applicability domain (AD) of the LR models was checked; it is noteworthy to mention that the test set compounds are within the AD of the developed QSTR model based on the standardization approach. We have also checked the AD for the test set compounds based on the developed PLS model using the DModX approach. It was found that all the test set compounds are within the AD (D-critical = 1.897) (Figure S2). The various descriptors obtained in the PLS equation are already elaborately explained with the probable mode of action towards cytotoxicity to the Chinese Hamster ovary cells.

The detailed mechanisms of the toxicity in terms of the descriptors are depicted in Figure 3.

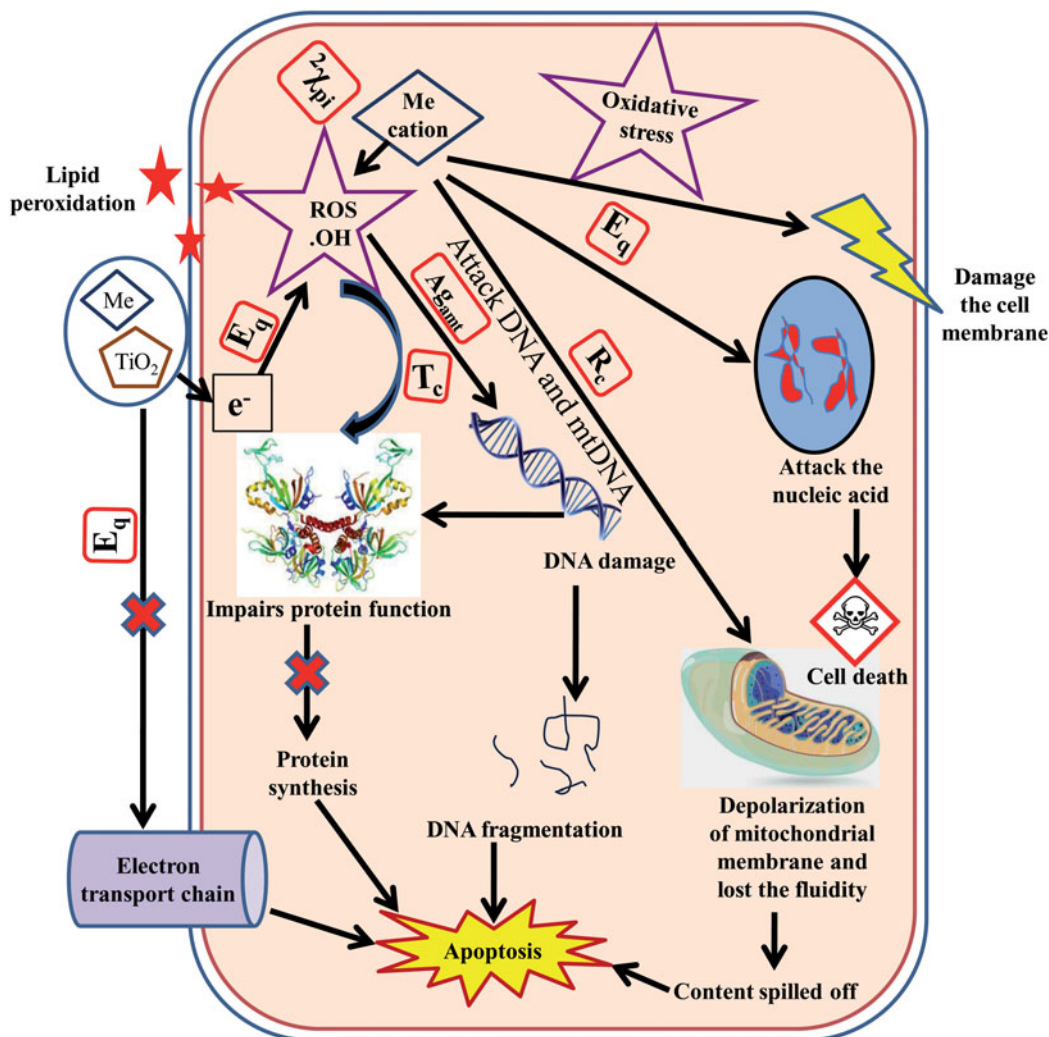


Figure 3. Mechanism of cytotoxicity of modified TiO_2 NPs towards the cell.

4. Conclusions

Nanotechnology has a very important impact in our daily life by giving useful solution to many global problems. The influence of nanotechnology is not fully established yet. According to the reviews of recently published papers (Gajewicz et al. 2012), there are still many gaps in the experimental data devoted to risk assessment of the nanoparticle available in today's market. The application of the theoretical methods is still in its developing stage whereas the usefulness of nanoparticle is rising day by day. In this context, we have developed interpretable QSTR models and predicted the cytotoxicity of modified TiO_2 based nanoparticle towards the Chinese hamster ovary cells using simple and easily calculated periodic table descriptors and examined the applicability of such descriptors to model metal oxide nanoparticles like any other

computational approaches. One of the aims of this work is to establish the simple periodic table descriptors useful for the modified nanoparticles for future use. It is believed that this type of descriptors can be used to develop the QSTR or QSAR models for other inorganic compounds also. The periodic table descriptors such as electrochemical equivalent (E_q), 2nd ionization potential (${}^2\chi_{pi}$), covalent radius (R_c), amount of Ag (Ag_{amt}) and thermal conductivity (T_c) can well explain the cytotoxicity without any exhaustive calculation, and thus it brings simplicity to the presented work. All the descriptors positively contributed to the response which means that increasing the descriptor values will also increase the cytotoxicity. Oxidation number, electronegativity, molecular weight as suggested by the various periodic table based descriptors play an important role in the

cytotoxicity of TiO₂ based NPs coated with various metal or mixture of metals in different concentrations. The transfer of electron from the valence band to the conduction band, the detachment of the metal cation from the surface of the modified metal oxide surface, increase of amount of silver ion in the cell and production of radicals due to lower oxidation number induce oxidative stress, depolarization of the mitochondria impair protein function, cause fragmentation of the DNA and thus cause apoptosis and death of the cell. The major finding of the work can be summarized below:

1. *Simplicity of the proposed models:* The models are developed with the additive mixture based descriptors, which is a relatively new concept to characterize and encode the modified heterogeneous nanoparticles. The advantage of this type of descriptors is that they allow the description of the heterogeneous nanoparticles taking into account the modification on the NPs as an example of the various amount of metal on the surface of TiO₂ NPs, and the variety of concentration of single metal clusters are effective in the calculation of the additive descriptors and interpretative nano-QSTR. Hence, the proposed models and approach have practical significance. The calculation of the periodic table descriptor is also not computationally demanding, and they can be easily obtained without any quantum chemical background.
2. *Mechanistic approach:* All the descriptors (E_q , χ_p , R_c , Ag_{amt} and T_c) are important for the cytotoxicity of the Chinese hamster ovary cell. Metals with high electronegativity, low oxidation state, tendency to lose electrons and easy detachment of the metal cations from the modified metal oxide surface may contribute to the cell toxicity. The success of this work is to use simple descriptors for the prediction of the cytotoxicity of the modified metal oxide with the probable mechanistic interpretation.
3. *Cost effective and time effective:* These simple descriptors as used in this study do not involve any hard or laborious calculation thus making the use of such descriptors simple and easy. For the calculation of descriptors, there is no need to use any computational software, only

the knowledge of periodic table is enough, making it both cost and time effective.

The periodic table based descriptors may thus be used to calculate the toxicity of any type of metal oxides in future studies also.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Financial assistance from the AICTE, New Delhi in the form of a fellowship to JR is thankfully acknowledged. PKO thanks the UGC, New Delhi for financial assistance in the form of a fellowship (Letter number and date: F./PDFSS-2015-17-WES-11996; dated: 06/04/2016). KR thanks CSIR, New Delhi for financial assistance under a Major Research project (CSIR Project No. 01(2895)/17/EMR-II).

ORCID

Kunal Roy  <http://orcid.org/0000-0003-4486-8074>

References

- Altenburger, R., H. Walter, and M. Grote. 2004. "What Contributes to the Combined Effect of a Complex Mixture?" *Environmental Science and Technology* 38 (23): 6353–6362. doi:10.1021/es049528k.
- Ames, B. N., M. K. Shigenaga, and T. M. Hagen. 1993. "Oxidants, Antioxidants, and the Degenerative Diseases of Aging." *Proceedings of the National Academy of Sciences of the United States of America* 90 (17): 7915–7922. <https://doi.org/10.1073/pnas.90.17.7915>.
- Artiles, M. S., C. S. Rout, and T. S. Fisher. 2011. "Graphene-Based Hybrid Materials and Devices for Biosensing." *Advanced Drug Delivery Reviews* 63 (14–15): 1352–1360. doi:10.1016/j.addr.2011.07.005.
- Benigni, R., and A. Giuliani. 2003. "Putting the Predictive Toxicology Challenge into Perspective: Reflections on the Results." *Bioinformatics* 19 (10): 1194–1200. doi:10.1093/bioinformatics/btg099.
- Bilberg, K., K. B. Døving, K. Beedholm, and E. Baatrup. 2011. "Silver Nanoparticles Disrupt Olfaction in Crucian Carp (*Carassius Carassius*) and Eurasian Perch (*Perca Fluviatilis*)." *Aquatic Toxicology* 104 (1–2): 145–152. doi:10.1016/j.aquatox.2011.04.010.
- Burello, E., A. Worth. 2012. "Development and evaluation of structure–reactivity models for predicting the in vitro oxidative stress of metal oxide nanoparticles." Towards Efficient Designing of Safe Nanomaterials: Innovative Merge of Computational Approaches and Experimental

- Techniques Cambridge, United Kingdom: Royal Society of Chemistry: 257–283.
- Cattaneo, A. G., Gornati, R. E. Sabbioni, M. Chiriva, -Internati, E. Cobos, M. R. Jenkins, and G. Bernardini. 2010. "Nanotechnology and Human Health: risks and Benefits." *Journal of Applied Toxicology: Jat* 30 (8): 730–744. doi: [10.1002/jat.1609](https://doi.org/10.1002/jat.1609).
- Cheng, X., W. Zhang, Y. Ji, J. Meng, H. Guo, J. Liu, X. Wu, and H. Xu. 2013. "Revealing Silver Cytotoxicity Using Au Nanorods/Ag Shell Nanostructures: disrupting Cell Membrane and Causing Apoptosis through Oxidative Damage." *RSC Advances* 3 (7): 2296–2305. doi:[10.1039/c2ra23131j](https://doi.org/10.1039/c2ra23131j).
- Chin, W. W. 1998. "The Partial Least Squares Approach to Structural Equation Modeling." *Modern Methods for Business Research* 295 (2): 295–336.
- Chirico, N., and P. Gramatica. 2011. "Real External Predictivity of QSAR Models: How to Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient." *Journal of Chemical Information and Modeling* 51 (9): 2320–2335. doi:[10.1021/ci200211n](https://doi.org/10.1021/ci200211n).
- Cho, M., H. Chung, W. Choi, and J. Yoon. 2004. "Linear Correlation between Inactivation of *E. coli* and OH Radical Concentration in TiO₂ Photocatalytic Disinfection." *Water Research* 38 (4): 1069–1077. doi:[10.1016/j.watres.2003.10.029](https://doi.org/10.1016/j.watres.2003.10.029).
- Concu, R., V. V. Kleandrova, A. Speck-Planche, and M. N. D. S. Cordeiro. 2017. "Probing the Toxicity of Nanoparticles: A Unified in Silico Machine Learning Model Based on Perturbation Theory." *Nanotoxicology* 11 (7): 891–906. doi: [10.1080/17435390.2017.1379567](https://doi.org/10.1080/17435390.2017.1379567).
- De, P., S. Kar, K. Roy, and J. Leszczynski. 2018. "Second Generation Periodic Table-Based Descriptors to Encode Toxicity of Metal Oxide Nanoparticles to Multiple Species: QSTR Modeling for Exploration of Toxicity Mechanisms." *Environmental Science: Nano* 5 (11): 2742–2760. doi: [10.1039/C8EN00809D](https://doi.org/10.1039/C8EN00809D).
- Drake, P. L., and K. J. Hazelwood. 2005. "Exposure-Related Health Effects of Silver and Silver Compounds: A Review." *The Annals of Occupational Hygiene* 49 (7): 575–585. doi: [10.1093/annhyg/mei019](https://doi.org/10.1093/annhyg/mei019).
- Dreher, K. L. 2004. "Health and Environmental Impact of Nanotechnology: Toxicological Assessment of Manufactured Nanoparticles". *Toxicology Sciences* 77 (1): 3–5. doi:[10.1093/toxsci/kfh041](https://doi.org/10.1093/toxsci/kfh041).
- ECHA 2012. Experts Workshop on Read-Across Assessment Organised by ECHA with the active support from Cefic-LRI (October 3, 2012). Accessed on December 2018. http://cefic-lri.org/wp-content/uploads/2014/03/ECHA-Cefic-LRI-Read-across-Workshop-Report_171211-FINAL.pdf
- European, C. 2006. "Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), Establishing a European Chemicals Agency, Amending Directive 1999/45/EC and Repealing Council Regulation (EEC) No 793/93 and Commission Regulation (EC) No 1488/94 as Well as Council Directive 76/769/EEC and Commission Directives 91/155/EEC, 93/67/EEC, 93/105/EC and 2000/21/EC." *Official Journal European Union* 396: 374–375. <https://eur-lex.europa.eu/eli/reg/2006/1907/2014-04-10>
- Foldbjerg, R., D. A. Dang, and H. Autrup. 2011. "Cytotoxicity and Genotoxicity of Silver Nanoparticles in the Human Lung Cancer Cell Line, A549." *Archives of Toxicology* 85 (7): 743–750. doi:[10.1007/s00204-010-0545-5](https://doi.org/10.1007/s00204-010-0545-5).
- Gajewicz, A., B. Rasulev, T. C. Dinadayalane, P. Urbaszek, T. Puzyn, D. Leszczynska, and J. Leszczynski. 2012. "Advancing Risk Assessment of Engineered Nanomaterials: Application of Computational Approaches." *Advanced Drug Delivery Reviews* 64 (15): 1663–1693. doi:[10.1016/j.addr.2012.05.014](https://doi.org/10.1016/j.addr.2012.05.014).
- Handy, R. D., R. Owen, and E. Valsami-Jones. 2008. "The Ecotoxicology of Nanoparticles and Nanomaterials: current Status, Knowledge Gaps, Challenges, and Future Needs." *Ecotoxicology* 17 (5): 315–325. doi:[10.1007/s10646-008-0206-0](https://doi.org/10.1007/s10646-008-0206-0).
- Handy, R. D., F. Von der Kammer, J. R. Lead, M. Hassellöv, R. Owen, and M. Crane. 2008. "The Ecotoxicology and Chemistry of Manufactured Nanoparticles." *Ecotoxicology* 17 (4): 287–314. doi:[10.1007/s10646-008-0199-8](https://doi.org/10.1007/s10646-008-0199-8).
- Hansen, S. F., A. Maynard, A. Baun, and J. A. Tickner. 2008. "Late Lessons from Early Warnings for Nanotechnology." *Nature Nanotechnology* 3 (8): 444. doi:[10.1038/nnano.2008.198](https://doi.org/10.1038/nnano.2008.198).
- Hewlett, P. S., and C. F. Wilkinson. 1967. "Quantitative Aspects of the Synergism between Carbaryl and Some 1, 3-Benzodioxole (Methylenedioxyphenyl) Compounds in Housefly." *Journal of the Science of Food and Agriculture* 18 (7): 279–282. doi:[10.1002/jfsa.2740180703](https://doi.org/10.1002/jfsa.2740180703).
- Highsmith, J. 2014. "Nanoparticles in Biotechnology, Drug Development and Drug Delivery." Report BIO113B. Wellesley, MA: BCC Research. Available online: <http://bccresearch.com/marketresearch/biotechnology/nanoparticles-biotechnology-drug-development-drug-delivery-report-bio113b.html>
- Islam, N., and K. Miyazaki. 2010. "An Empirical Analysis of Nanotechnology Research Domains." *Technovation* 30 (4): 229–237. doi:[10.1016/j.technovation.2009.10.002](https://doi.org/10.1016/j.technovation.2009.10.002).
- Jiang, X., R. Foldbjerg, T. Miclaus, L. Wang, R. Singh, Y. Hayashi, D. Sutherland, C. Chen, H. Autrup, and C. Beer. 2013. "Multi-Platform Genotoxicity Analysis of Silver Nanoparticles in the Model Cell Line CHO-K1." *Toxicology Letters* 222 (1): 55–63. doi:[10.1016/j.toxlet.2013.07.011](https://doi.org/10.1016/j.toxlet.2013.07.011).
- Kang, S., M. Pinault, L. D. Pfeifferle, and M. Elimelech. 2007. "Single-Walled Carbon Nanotubes Exhibit Strong Antimicrobial Activity." *Langmuir* 23 (17): 8670–8673. doi: [10.1021/la701067r](https://doi.org/10.1021/la701067r).
- Kar, S., A. Gajewicz, T. Puzyn, K. Roy, and J. Leszczynski. 2014. "Periodic Table-Based Descriptors to Encode Cytotoxicity Profile of Metal Oxide Nanoparticles: A Mechanistic QSTR Approach." *Ecotoxicology and Environmental Safety*.107: 162–169. doi:[10.1016/j.ecoenv.2014.05.026](https://doi.org/10.1016/j.ecoenv.2014.05.026).

- Kar, S., A. Gajewicz, K. Roy, J. Leszczynski, and T. Puzyn. 2016. "Extrapolating between Toxicity Endpoints of Metal Oxide Nanoparticles: Predicting Toxicity to Escherichia coli and Human Keratinocyte Cell Line (HaCaT) with Nano-QTR." *Ecotoxicology and Environmental Safety* 126: 238–244. doi:10.1016/j.ecoenv.2015.12.033.
- Kim, Y.-J., S. I. Yang, and J.-C. Ryu. 2010. "Cytotoxicity and Genotoxicity of Nano-Silver in Mammalian Cell Lines." *Molecular & Cellular Toxicology* 6 (2): 119–125. doi: 10.1007/s13273-010-0018-1.
- Kittler, S., C. Greulich, J. Diendorf, M. Köllner, and M. Eppler. 2010. "Toxicity of Silver Nanoparticles Increases during Storage Because of Slow Dissolution under Release of Silver Ions." *Chemistry of Materials* 22 (16): 4548–4554. doi: 10.1021/cm100023p.
- Kleandrova, V. V., F. Luan, H. González-Díaz, J. M. Ruso, A. Melo, A. Speck-Planche, and M. N. L. D. S. Cordeiro. 2014. "Computational Ecotoxicology: Simultaneous Prediction of Ecotoxic Effects of Nanoparticles under Different Experimental Conditions." *Environment International* 73: 288–294. doi:10.1016/j.envint.2014.08.009.
- Kleandrova, V. V., F. Luan, H. González-Díaz, J. M. Ruso, A. Speck-Planche, and M. N. D. S. Cordeiro. 2014. "Computational Tool for Risk Assessment of Nanomaterials: Novel QSTR-Perturbation Model for Simultaneous Prediction of Ecotoxicity and Cytotoxicity of Uncoated and Coated Nanoparticles under Multiple Experimental Conditions." *Environmental Science and Technology* 48 (24): 14686–14694. doi:10.1021/es503861x.
- Koppenol, W. H. 2001. "The Haber-Weiss cycle-70 years later." *REDOX Report: Communications in Free Radical Research* 6 (4): 229–234. doi:10.1179/135100001101536373.
- Lin, W., Y. W. Huang, X. D. Zhou, and Y. Ma. 2006a. "In vitro toxicity of silica nanoparticles in human lung cancer cells." *Toxicology and Applied Pharmacology* 217 (3): 252–259. doi:10.1016/j.taap.2006.10.004.
- Lin, W., Y.-W. Huang, X.-D. Zhou, and Y. Ma. 2006b. "Toxicity of Cerium Oxide Nanoparticles in Human Lung Cancer Cells." *International Journal of Toxicology* 25 (6): 451–457. doi:10.1080/10915810600959543.
- Lovrić, J., S. J. Cho, F. M. Winnik, and D. Maysinger. 2005. "Unmodified Cadmium Telluride Quantum Dots Induce Reactive Oxygen Species Formation Leading to Multiple Organelle Damage and Cell Death." *Chemistry and Biology* 12 (11): 1227–1234. doi:10.1016/j.chembiol.2005.09.008.
- Luan, F., V. V. Kleandrova, H. González-Díaz, J. M. Ruso, A. Melo, A. Speck-Planche, and M. N. D. S. Cordeiro. 2014. "Computer-Aided Nanotoxicology: assessing Cytotoxicity of Nanoparticles under Diverse Experimental Conditions by Using a Novel QSTR-Perturbation Approach." *Nanoscale* 6 (18): 10623–10630. doi:10.1039/C4NR01285B.
- Lubick, N. 2008. Nanosilver Toxicity: ions, Nanoparticles- or Both? *Environ Sci Technol* 42(23), 8617–8617. doi: 10.1021/es8026314.
- Melagraki, G., and A. Afantitis. 2014. "Enalos InSilicoNano Platform: An Online Decision Support Tool for the Design and Virtual Screening of Nanoparticles." *RSC Advances* 4 (92): 50713–50725. doi:10.1039/C4RA07756C.
- Mikolajczyk, A., A. Gajewicz, E. Mulkiewicz, B. Rasulev, M. Marchelek, M. Diak, S. Hirano, A. Zaleska-Medynska, and T. Puzyn. 2018. "Nano-QSAR Modeling for Ecosafe Design of Heterogeneous TiO₂-Based Nano-Photocatalysts." *Environmental Science: Nano* 5 (5): 1150–1160. doi: 10.1039/C8EN00085A.
- Mikolajczyk, A., A. Malankowska, G. Nowaczyk, A. Gajewicz, S. Hirano, S. Jurga, A. Zaleska-Medynska, and T. Puzyn. 2016. "Combined Experimental and Computational Approach to Developing Efficient Photocatalysts Based on Au/Pd–TiO₂ Nanoparticles." *Environmental Science: Nano* 3 (6): 1425–1435. doi:10.1039/C6EN00232C.
- Morris, J., J. Willis, D. De Martinis, B. Hansen, H. Laursen, J. R. Sintes, P. Kearns, and M. Gonzalez. 2011. "Science Policy Considerations for Responsible Nanotechnology Decisions." *Nature Nanotechnology* 6 (2): 73. doi:10.1038/nnano.2010.191.
- National Research Council. 2002. *Small Wonders, Endless Frontiers: A Review of the National Nanotechnology Initiative*. Washington, DC: National Academies Press.
- Neal, A. L. 2008. "What Can Be Inferred from Bacterium–Nanoparticle Interactions about the Potential Consequences of Environmental Exposure to Nanoparticles." *Ecotoxicology* 17 (5):362. doi:10.1007/s10646-008-0217-x.
- OECD 2014. *Guidance on Grouping of Chemicals, Series on Testing and Assessment*. No. 194 2nd ed. Organisation of Economic Cooperation and Development: Paris, France.
- Ojha, P. K., I. Mitra, R. N. Das, and K. Roy. 2011. "Further Exploring r_m^2 Metrics for Validation of QSPR Models." *Chemometrics and Intelligent Laboratory Systems* 107 (1): 194–205. doi:10.1016/j.chemolab.2011.03.011.
- Ozben, T. 2007. "Oxidative Stress and Apoptosis: impact on Cancer Therapy." *Journal of Pharmaceutical Sciences* 96 (9): 2181–2196. doi:10.1002/jps.20874.
- Plackett, R. L., and P. S. Hewlett. 1967. "A Comparison of Two Approaches to the Construction of Models for Quantal Responses to Mixtures of Drugs." *Biometrics* 23 (1): 27–44. doi:10.2307/2528279.
- Puzyn, T., A. Gajewicz, D. Leszczynska, and J. Leszczynski. 2010. *Nanomaterials—the Next Great Challenge for QSAR Modelers. Recent Advances in QSAR Studies*. Dordrecht: Springer; 383–409.
- Puzyn, T., J. Leszczynski, and M. T. Cronin. 2010. *Recent Advances in QSAR Studies: methods and Applications*. The Netherlands: Springer Science & Business Media.
- Puzyn, T., B. Rasulev, A. Gajewicz, X. Hu, T. P. Dasari, A. Michalkova, H.-M. Hwang, A. Toropov, D. Leszczynska, and J. Leszczynski. 2011. "Using nano-QSAR to Predict the Cytotoxicity of Metal Oxide Nanoparticles." *Nature Nanotechnology* 6 (3): 175. doi:10.1038/nnano.2011.10.
- RCC-NI 2011. Regulatory Cooperation Council Nanotechnology Initiative Work Element 4 Final Report Assessment of Nanomaterial Uses in Canada and the US. <http://www.ec.gc.ca/scitech/default.asp?lang=En&n=6A2D63E5-1&xsl=privateArticles2,viewfull&po=08C45DB6>.

- Roy, K., and P. Ambure. 2016. "The 'Double Cross-Validation' Software Tool for MLR QSAR Model Development." *Chemometrics and Intelligent Laboratory Systems* 159: 108–126. doi:10.1016/j.chemolab.2016.10.009.
- Roy, K., and R. N. Das. 2013. "QSTR with Extended Topochemical Atom (ETA) Indices. 16. Development of Predictive Classification and Regression Models for Toxicity of Ionic Liquids towards Daphnia Magna." *Journal of hazardous materials* 254: 166–178. doi:10.1016/j.jhazmat.2013.03.023.
- Roy, K., R. N. Das, P. Ambure, and R. B. Aher. 2016. "Be Aware of Error Measures. Further Studies on Validation of Predictive QSAR Models." *Chemometrics and Intelligent Laboratory Systems* 152: 18–33. doi:10.1016/j.chemolab.2016.01.008.
- Roy, K., S. Kar, and P. Ambure. 2015. "On a Simple Approach for Determining Applicability Domain of QSAR Models." *Chemometrics and Intelligent Laboratory Systems* 145: 22–29. doi:10.1016/j.chemolab.2015.04.013.
- Sawai, J., H. Kojima, H. Igarashi, A. Hashimoto, S. Shoji, T. Sawaki, A. Hakoda, E. Kawada, T. Kokugan, and M. Shimizu. 2000. "Antibacterial Characteristics of Magnesium Oxide Powder." *World Journal of Microbiology and Biotechnology* 16 (2): 187–194. doi:10.1023/A:1008916209784.
- Serpone, N., and A. V. Emeline. 2012. Semiconductor Photocatalysis—, Past, Present, and Future Outlook. *J Phys Chem Lett* 3(5), 673–677. doi: 10.1021/jz300071j.
- Speck-Planche, A., V. V. Kleandrova, F. Luan, and M. N. Ds Cordeiro. 2015. "Computational Modeling in Nanomedicine: prediction of Multiple Antibacterial Profiles of Nanoparticles Using a Quantitative Structure- Activity Relationship Perturbation Model." *Nanomedicine* 10 (2): 193–204. doi:10.2217/nnm.14.96.
- Todeschini, R., and V. Consonni. 2009. *Molecular Descriptors for Chemoinformatics*, I and II. Weinheim: John Wiley & Sons.
- Tropsha, A. 2003. In: Burger's Medicinal Chemistry and Drug Discovery, edited by Donald J. Abraham, John Wiley & Sons, NY, Chapter 2. <https://doi.org/10.1002/0471266949.bmc002.pub2> Recent advances in development, validation, and exploitation of QSAR models.
- Tropsha, A. 2010. "Best Practices for QSAR Model Development, Validation, and Exploitation." *Molecular Informatics* 29 (6–7): 476–488. doi:10.1002/minf.201000061.
- Umetrics, M. 2013. *User Guide to SIMCA*. Malmö(Sweden): MKS Umetrics AB.
- Walker, J. D., M. Enache, and J. C. Dearden. 2003. "Quantitative Cationic-Activity Relationships for Predicting Toxicity of Metals." *Environmental Toxicology and Chemistry* 22 (8): 1916–1935. doi:10.1897/02-568.
- Wang, D., Y. Gao, Z. Lin, Z. Yao, and W. Zhang. 2014. "The Joint Effects on Photobacterium phosphoreum of Metal Oxide Nanoparticles and Their Most Likely Coexisting Chemicals in the Environment." *Aquatic Toxicology* 154: 200–206. doi:10.1016/j.aquatox.2014.05.023.
- Xia, T., M. Kovochich, J. Brant, M. Hotze, J. Sempf, T. Oberley, C. Sioutas, J. I. Yeh, M. R. Wiesner, and A. E. Nel. 2006. "Comparison of the Abilities of Ambient and Manufactured Nanoparticles to Induce Cellular Toxicity according to an Oxidative Stress Paradigm." *Nano Letters* 6 (8): 1794–1807. doi:10.1021/nl061025k.
- Xu, S., and N. Nirmalakhandan. 1998. "Use of QSAR Models in Predicting Joint Effects in Multi-Component Mixtures of Organic Chemicals." *Water Research* 32 (8): 2391–2399. doi:10.1016/S0043-1354(98)00006-2.
- Zhang, L., P-J Zhou, F. Yang, and Z-D Wang. 2007. "Computer-Based QSARs for Predicting Mixture Toxicity of Benzene and Its Derivatives." *Chemosphere* 67 (2): 396–401. doi:10.1016/j.chemosphere.2006.09.018.
- Zhang, H., Z. Ji, T. Xia, H. Meng, C. Low-Kam, R. Liu, S. Pokhrel, S. Lin, X. Wang, Y. Liao, M. Wang, L. Li, R. Rallo, R. Damoiseaux, D. Telesca, L. Madler, Y. Cohen, J. I. Zink, A. E. Nel. 2012. "Use of Metal Oxide Nanoparticle Band Gap to Develop a Predictive Paradigm for Acute Pulmonary Inflammation based on Oxidative Stress." *ACS Nano* 6: 4349–4368. doi:10.1021/nn3010087.
- Zhu, H., T. M. Martin, L. Ye, A. Sedykh, D. M. Young, and A. Tropsha. 2009. "Quantitative Structure – Activity Relationship Modeling of Rat Acute Toxicity by Oral Exposure." *Chemical Research in Toxicology* 22 (12): 1913–1921. doi:10.1021/tx900189p.