“Quantitative Structure-Activity Relationship (QSAR) models”

Ceyda Oksel
Xue Z. Wang

University of Leeds

27.03.2013
Purpose of Today

- Theory of QSAR
- New Data Mining Methods
- Nanomaterial Characterization
- Nanotoxicity
- Application of QSAR model in nanotoxicity modeling
Part I

Theory of QSAR
Crum Brown and Fraser\textsuperscript{[1]} expressed the idea that there was a relationship between activity and chemical structure.

Richet\textsuperscript{[2]} correlated toxicities of simple organic molecules with their solubility in water.

Meyer\textsuperscript{[3]} and Overton\textsuperscript{[4]} found linear relationships between the toxicity of organic compounds and their lipophilicity.

Hansch\textsuperscript{[5]} published a free-energy related model to correlate biological activities with physicochemical Properties.

The father of the concept of quantitative structure-activity relationship (QSAR), the quantitative correlation of the physicochemical properties of molecules with their biological activities\textsuperscript{[5]}

A QSAR is a statistical model that relates a set of structural descriptors of a chemical compound to its biological activity.

The presence of particular characteristics increase the probability that the compound is toxic.
Nano-QSAR

QSAR models are very useful in case of the classic chemicals but the concept of nano-QSAR is still under development.

QSAR has been widely used to predict the toxicity of substances in bulk form (especially drug-like compounds) but, up to date, QSAR studies for the prediction of nanoparticle toxicity have been rarely reported.

The European system for new chemical management (REACH) promotes QSAR methods as an alternative way of toxicity testing.
Why do we need QSAR models?

- All chemical substances need to be tested in terms of their toxicological and environmental properties before their use.
- There are several reasons to use QSAR Models: very fast, often free, reduce the number of animals used in experiments.

Alternative, fast developing applications

The most important sources of information

Reduction in the time, cost and animal testing

QSAR: quantitative structure-activity relationship

## Purpose of QSAR

<table>
<thead>
<tr>
<th><strong>Purpose</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>To predict biological activity and physico-chemical properties by rational means</td>
<td></td>
</tr>
<tr>
<td>To comprehend and rationalize the mechanisms of action within a series of chemicals</td>
<td></td>
</tr>
<tr>
<td>Savings in the cost of product development</td>
<td></td>
</tr>
<tr>
<td>Predictions could reduce the requirement for lengthy and expensive animal tests.</td>
<td></td>
</tr>
<tr>
<td>Reduction of animal tests, thus reducing animal use and obviously pain and discomfort to animals</td>
<td></td>
</tr>
<tr>
<td>Other areas of promoting green and greener chemistry to increase efficiency and eliminate waste</td>
<td></td>
</tr>
</tbody>
</table>
Requirements for QSAR

1. Set of molecules
2. Set of molecular descriptors
3. Measured biological activity or property
4. Statistical methods

How many compounds are required to develop a QSAR?

There is no direct and simple answer → “as many as possible”

To provide some guide, it is widely accepted that between five and ten compounds are required for every descriptor in a QSAR [1]

There are a large number of applications of these models within industry, academia and governmental (regulatory) agencies.

- The estimation of physico-chemical properties, biological activities and understanding the physicochemical features behind a biological response in drug design.
- The rational design of numerous other products such as surface-active agents, perfumes, dyes, and fine chemicals.
- The prediction of a variety of physico-chemical properties of molecules.
- The prediction of fate of molecules which are released into the environment.
- The identification of hazardous compounds at early stages of product development, the prediction of toxicity to humans and environment.
What is required for a good QSAR Model?

The QSAR model should meet the requirements of the OECD principles [1]:

1. a defined endpoint;
2. an unambiguous algorithm;
3. a defined domain of applicability;
4. appropriate measures of goodness-of-fit, robustness and predictivity;
5. a mechanistic interpretation if possible.

Common QSAR Modelling Errors

- Uninformative descriptors
- Poor descriptor selection and chance correlations
- Modelling complex, nonlinear structure property relationships with linear models
- Incorrectly validating QSPR models
- Not understanding the domain of applicability of models

Methods

• QSAR Modelling process consists of 5 main steps.

Main Steps

Molecular Structure

Representation

Molecular Descriptors

Predictive Model

Model Validation

Applicability Domain of QSAR

• Begins with the selection of molecules to be used
• Selection of descriptors; numerical representer of molecular features (e.g. number of carbon)
• Original descriptor pool must be reduced in size
• Model building
• The reliability of the model should be tested

Types of Molecular Descriptors
(topological, geometric, electronic and hybrid)

1D Descriptors
2D Descriptors
3D Descriptors
Molecular Descriptors\textsuperscript{[1]}

are numerical values that characterize properties of molecules.


Different Statistical Methods have been used in QSAR for the extraction of useful information from the data.

**Regression Problems**
- Multiple linear regression
- Partial least squares
- Feed-forward back propagation neural network
- General regression neural network

**Classification Problems**
- Linear Discriminant analysis
- Logistic regression
- Decision tree
- K-nearest neighbor
- Probabilistic Neural Network
- Support vector machine

**Recently-developed Methods**
- Kernel partial least square
- Robust continuum regression
- Local lazy regression
- Fuzzy interval number k-nearest neighbor
- Fast projection plane classifier

Data Mining Methods

• a relatively new field which share some techniques with traditional data analysis but also brings a lot of new visions and techniques.

‘DM is the discovery of useful Information and Knowledge from a large collection of Data, in order to help the Decision making’

- **Data**: records of numerical data, symbols, images, documents
- **Knowledge**: Rules: IF .. THEN ..
  Cause-effect relationships
  Decision trees
  Patterns: abnormal, normal operation
  Predictive equations
Data mining shares many techniques and tools with traditional data analysis (such as the ones listed below). But it emphasises the analysis of very large databases, and can deal with complicated cases which cannot be handled by traditional methods.

- Clustering
- Classification
- Conceptual Clustering
- Inductive learning
- Summarisation
- Regression
- Case-based Learning

Data mining research has also generated some new techniques that were not seen in traditional data analysis, such as **Link Analysis**
Example: Ten data cases, three variables, $x1$, $x2$, and $x3$

How can we automatically uncover the dependency relationship between them?

<table>
<thead>
<tr>
<th>DATA SET</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$x1$</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$x2$</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$x3$</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

WHICH RELATIONSHIP?

This relationship means, $x2$ is dependent on $x1$, and $x3$ is dependent on $x2$, but $x1$ is independent of $x2$ and $x3$.

This relationship means, $x2$ is dependent on $x1$, $x3$ is also dependent on $x1$, but $x2$ and $x3$ are independent of each other.
New DM Technique Example 2: Inductive Data Mining for Automatic Generation of Decision Trees from Data

<table>
<thead>
<tr>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>y</th>
<th>x1</th>
<th>x2</th>
<th>x3</th>
<th>x4</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>55.33</td>
<td>1.72</td>
<td>54</td>
<td>1.66219</td>
<td>92.19</td>
<td>71.31</td>
<td>3.44</td>
<td>55</td>
<td>1.60325</td>
<td>90.51</td>
</tr>
<tr>
<td>59.13</td>
<td>1.2</td>
<td>53</td>
<td>1.58399</td>
<td>92.74</td>
<td>72.3</td>
<td>4.02</td>
<td>55</td>
<td>1.66783</td>
<td>90.24</td>
</tr>
<tr>
<td>57.39</td>
<td>1.42</td>
<td>55</td>
<td>1.61731</td>
<td>91.88</td>
<td>68.81</td>
<td>6.88</td>
<td>55</td>
<td>1.69836</td>
<td>91.01</td>
</tr>
<tr>
<td>56.43</td>
<td>1.78</td>
<td>55</td>
<td>1.66228</td>
<td>92.8</td>
<td>66.61</td>
<td>2.31</td>
<td>52</td>
<td>1.77967</td>
<td>91.9</td>
</tr>
<tr>
<td>55.98</td>
<td>1.58</td>
<td>54</td>
<td>1.63195</td>
<td>92.56</td>
<td>63.66</td>
<td>2.99</td>
<td>52</td>
<td>1.81271</td>
<td>91.92</td>
</tr>
<tr>
<td>56.16</td>
<td>2.12</td>
<td>56</td>
<td>1.68034</td>
<td>92.61</td>
<td>63.85</td>
<td>0.24</td>
<td>50</td>
<td>1.81485</td>
<td>92.16</td>
</tr>
<tr>
<td>54.85</td>
<td>1.17</td>
<td>54</td>
<td>1.58206</td>
<td>92.33</td>
<td>67.25</td>
<td>0</td>
<td>53</td>
<td>1.72526</td>
<td>91.36</td>
</tr>
<tr>
<td>52.83</td>
<td>1.5</td>
<td>58</td>
<td>1.54998</td>
<td>92.22</td>
<td>67.19</td>
<td>0</td>
<td>52</td>
<td>1.86782</td>
<td>92.16</td>
</tr>
</tbody>
</table>

x1, x2, x3 are different measures of feed composition
x4 is the log of a combination of operating conditions
y – product quality measure
y = ‘good’ if > 92,
   = ‘low’ if between 91-92,
   = ‘very low’ if < 91

**Aim:** to find which feed composition or operational condition variables have the most important impact on product quality
Automatically generated decision tree

Product quality is mainly determined by $x_1$ and $x_4$, almost has nothing to do with $x_2$ and $x_3$

- If $x_1 < 55.4$, the probability that product quality is good ($y > 92$) is high
- If $x_1 > 55.4$ but $x_4 > 1.8$, the probability that product quality is good is also high
- Other two regions, product quality is either low or very low

If $x_1 = [44.6, 55.4]$, then conditional probability of $y = 3$ is 0.9
If $x_4 = [1.8, 2.3]$, then the conditional probability of $y = 3$ is 1.0
(1) Generation of a population of solutions

(2) Repeat steps (i) and (ii) until the stop criteria are satisfied:

(i) calculate the fitness function values for each solution candidate

(ii) perform crossover and mutation to generate the next generation

(3) the best solution in all generations: the final solution

Automatic generation of decision tree from data
This is a new approach based on a genetic programming method
Two Data Sets

Data Set 1:
Concentration lethal to 50% of the population, LC50, 1/\log(\text{LC50}), of \textit{vibrio fischeri}, a bioluminescent bactorium
75 compounds 1069 molecular descriptors calculated by molecular modelling software DRAGON

Data Set 2:
Concentration effecting 50% of the population, EC50 of algae \textit{chlorella vulgaris}, by causing fluorescein diacetate to disappear
80 compounds 1150 descriptors
A method to classify toxicity endpoints into classes

Sample tree from Y-Adapt

Polar surface area ≤ 40.368

- Yes → Geary autocorrelation lag 4 weighted by atomic mass ≤ 0.007
  - Yes → Class 1 (21/21)
  - No → Geary autocorrelation lag 4 weighted by polarizability ≤ 1.181
    - Yes → Class 3 (9/10)
    - No → Broto-Moreau autocorrelation lag 2 weighted by polarizability ≤ 4.085
      - Yes → Class 2 (5/6)
      - No → Class 3 (6/6)

- No → H autocorrelation lag 0 weighted by Sanderson electro-negativities ≤ 2.111
  - Yes → Class 1 (8/8)
  - No → Class 2 (9/9)

A tree for the bacteria data using three classes.

For bacteria data: of 1069 descriptors, only 5 descriptors are important! Using the 5 descriptors, the tree model can predict with the accuracy > 96.7%
Y-Adapt for QSAR Toxicity Prediction: Data Set 1

A decision tree from generation 46 for the bacteria data, finding four classes from the continuous endpoint itself during induction.

Training accuracy: 93.3%
Test accuracy: 80.0%

1. Polar surface area ≤ 40.368
   - Yes: H autocorrelation lag 0 weighted by Sanderson electro-negativities ≤ 2.111
     - Yes: Class 1 (16/16)
     - No: Geary autocorrelation lag 4 weighted by atomic mass ≤ 0.007
       - Yes: Class 1 (13/13)
       - No: Class 2 (9/9)
   - No: 3D MoRSE signal 15 weighted 18 weighted by Sanderson electronegativites ≤ -0.644
     - Yes: Class 2 (5/7)
     - No: H autocorrelation lag 0 weighted by Sanderson electro-negativities ≤ 2.394
       - Yes: Class 3 (7/9)
       - No: Class 4 (6/6)

Class 1: Log(1/EC50) ≤ 4.08 29 cases
Class 2: 4.08 < Log(1/EC50) ≤ 4.51 16 cases
Class 3: 4.51 < Log(1/EC50) ≤ 4.99 9 cases
Class 4: otherwise 6 cases
**Y-Adapt for QSAR Toxicity Prediction: Data Set 2**

![Decision Tree](image)

**2nd dataset - algae data**

Of the 1150 descriptors, only 5 descriptors are important!

Training: 92.2%

Test: 81.3%
Different data mining techniques can be combined into a single system.
Model Validation

Model Applicability

Model Interpretation

Cross-validation (Q²)
• K-fold Cross Validation
• Leave-one-out Cross Validation

The statistical fit of model $R^2$

Standart Error of Prediction (S)

Model validation is required to ensure model reliability (quality of the model)
Ensures that the model is not due to chance factors!

The statistical fit of model is usually performed

Cross validation: dividing a data sample into subsets, performing the analysis on a single subset, using other subsets to confirm and validate the initial analysis

There is no agreed method for the validation of QSAR Models
QSAR Workflow[1]

• *Model applicability* allows us to decide whether the model will be useful for new data (correct use of the model).

• Finding the boundaries of the model to see how well it will work for other compounds.

• Not even a significant and validated QSAR model can be used for the entire universe of chemicals\(^1\)!

QSAR Tools for Toxicity Prediction

1. **TOPKAT** (by Komputer Assisted Technology): It uses QSARs to give numeric predictions, based on several descriptors.

2. **HazardExpert** (by CompuDrug): HazardExpert, produced by CompuDrug, is a rule-based system. The rules use toxicophores, or structural alerts derived from QSARs.

3. **CASE**: It can derive QSAR model from user’s data.

4. **ECOSAR**: It uses QSARs to predict the aquatic toxicity of chemicals for a variety of organisms.

5. **OASIS**: It uses hierarchical rules based on QSARs, which can be user derived.

6. **DEREK**: Deductive Estimation of Risk from Existing Knowledge (DEREK) is marketed by LHASA Ltd. It is an expert knowledge base system that predicts whether a chemical is toxic.

- Each of these QSAR tools have their own advantages and weaknesses for toxicity predictions.
- There are also several available software packages for descriptor calculations (e.g. DRAGON).
Challenges for QSAR [1]

1. If there is a measurement error in the experimental data, it is very likely that false correlations may arise.

2. If the training dataset is not large enough, the data collected may not reflect the complete property space. Consequently, many QSAR results cannot be used to confidently predict the most likely compounds of the best activity.

3. There are many successful applications of this method but one cannot expect the QSAR approach to work well all the time. **QSAR Models are easy to build but also very easy to get wrong.**

Part II

Nano-toxicity and Nanomaterial Characterization
Uncertainties at the early stages of a new technology bring new concerns!

- No standardized or validated methods for nanotoxicity testing
- Available toxicity data for NPs are inconsistent and confusing
- Blocks the development of ENPs risk reduction strategies

More than 800 nano-products

Macro concerns with nano-world!

• Safety to human health and environment
• Suitability of risk management strategies

Risk assessment for NMS [1]

- ENMs are already used in a number of commercial applications.
- Potential safety issues of NMs should be addressed at the same time as the tech. is developing

1. Hazard Identification
   • What dangers to human health may basically arise from the harmful agent?

2. Dose response relationship
   • What quantitative connections exist between the dose and the extent of the expected effect?

3. Exposure assessment
   • To what extent is the relevant population group exposed to the harmful agent?

4. Risk assessment
   • What is the nature and magnitude of the risk and how accurately can it be estimated?

Hazard Identification

Dose-response evaluation

Evaluation of toxicity

The evaluation of toxicity includes two steps:

- Hazard Identification
- Dose-response evaluation

Involves determining the conditions or levels at which the presence of a contaminant may trigger a biological response from the body.

Different amount of CNT dust exposure:

- At level of 0.1 mg/m$^3$: No effects were observed.
- At higher level of 0.5 mg/m$^3$: Side effects were detected in the lung.

The only reasonable way to obtain toxicity information for the large number of nanoparticles without testing every single one is to relate the physicochemical characteristics of NPs with their toxicity in a SAR (Structure Activity Relationship) or QSAR model.

Nanomaterial characterization is one of the first steps in toxicity test. However, it is a very challenging issue!!!
Possible toxicity-related Physico-chemical Properties

When a material is in a nano-sized form, its properties become different from the same material in a bulk form

- Fullerenes
- Carbon-based
- Carbon NTs
- Nanomaterials
- Graphene
- Different and Exceptional Properties

No certain information!!!

Recent studies showed that toxicity of NPs can be related to:

- Size and shape
- Size distribution
- Agglomeration state
- Porosity
- Structure-dependent elect.
- Electronic properties
- Surface area
- Surface chemistry
- Surface charge
- Crystal structure
- Composition
- Configuration
- Coating
- Aggregation state
- Metal content

POSSIBLE factors[1] inducing toxicity of chemicals

more research is required to reach a clear understanding

---

## Nanomaterial Characterization

Several different methods and instruments are used to determine the properties on NMs.

<table>
<thead>
<tr>
<th>Property</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape and aggregation state</td>
<td>High Resolution Microscopy</td>
</tr>
<tr>
<td>Composition, purity and surface chemistry</td>
<td>Spectroscopy and Chromatography Methods</td>
</tr>
<tr>
<td>Surface charge</td>
<td>Zeta Potential Analysis</td>
</tr>
<tr>
<td>Crystal structure</td>
<td>X-Ray Diffraction</td>
</tr>
<tr>
<td>Particle number and size distribution</td>
<td>Several Techniques (e.g. SEM, TEM, DLS)</td>
</tr>
<tr>
<td>Surface Area</td>
<td>Brunauer-Emmett-Teller Adsorption</td>
</tr>
</tbody>
</table>

There is no single instrument that is the right tool for every test. There are in fact more than 400 different techniques for particle counting and characterization.
Some NPs

Have a tendency to form clusters (e.g. aluminium oxide)

in an agglomeration state, some NPs behave like larger particles in the media

Toxicity may also depend on the size of agglomerate

NOT on original particle size itself !!!

If the NPs aggregate

How to characterize aggregation? Texture analysis?

### Challenges for nano-QSAR?

Table shows the results of 3 different toxicity studies for silver NMs:

<table>
<thead>
<tr>
<th>NP type</th>
<th>Size (nm)</th>
<th>Doses (mg/ml)</th>
<th>Cell</th>
<th>Method</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ag</td>
<td>7-20</td>
<td>0.8-50</td>
<td>Human fibrosorcoma and skin carcinoma</td>
<td>XTT assay</td>
<td>Non-toxic</td>
</tr>
<tr>
<td>Ag</td>
<td>&lt;15</td>
<td>0.7</td>
<td>Human hepatoma</td>
<td>MTT Alamar blue and LDH assay</td>
<td>Oxidative stress mediated toxicity</td>
</tr>
<tr>
<td>Ag</td>
<td>15 and 100</td>
<td>10-250</td>
<td>Rat liver delivered cell line</td>
<td>MTT, LDH, GSH level, ROS and MMP</td>
<td>Toxic</td>
</tr>
</tbody>
</table>

---

**Differences in NMs used (size)**

**Differences in experimental conditions**

strongly affects the results

---


NMs are problematic for QSAR modellers because of the several reasons:

1. Lack of high-quality experimental data
2. Lack of conceptual frameworks for grouping NPs according to mode of physicochemical properties and toxic action
3. Lack of sufficient molecular descriptors able to express specificity of “nano” structure
4. Lack of rational modeling procedures to screen large numbers of structurally diversified nanoparticles.
5. Limited knowledge on the interactions between NPs and biological systems

Part III

Case Study
despite all limitations...
The aim of this study was to test the hypothesis that nanoparticle toxicity is a function of some measurable physical characteristics and that Structure Activity Relationship (SAR) might ultimately be a useful tool for nanotoxicology. A panel of 18 nanoparticles was chosen to include carbon-based materials and metal oxides. Comparative toxicity of the nanoparticles in the panel was assessed, using a number of different measures of apoptosis and necrosis; haemolytic effects and the impact, on cell morphology were also assessed. Structural and composition properties were measured including size distribution, surface area, morphological parameters, metal concentration, reactivity, free radical generation and zeta potential. Toxicity end points were processed using principal component analysis and clustering algorithms.
# The Set of Eighteen NPs

<table>
<thead>
<tr>
<th>Particles</th>
<th>Qualifier</th>
<th>Number</th>
<th>Supplier</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon black</td>
<td>Printex 90</td>
<td>N1</td>
<td>Degussa</td>
<td>Dry</td>
</tr>
<tr>
<td>Diesel exhaust particles</td>
<td>EPA</td>
<td>N2</td>
<td>EPA</td>
<td>Dry</td>
</tr>
<tr>
<td>Nanotubes</td>
<td>Japanese</td>
<td>N3</td>
<td>Vicki Stone</td>
<td>Dry</td>
</tr>
<tr>
<td>Fullerene</td>
<td>Unmodified</td>
<td>N4</td>
<td>Sigma</td>
<td>Dry</td>
</tr>
<tr>
<td>Polystyrene latex</td>
<td>Unmodified</td>
<td>N5</td>
<td>Polysciences</td>
<td>Sus.</td>
</tr>
<tr>
<td></td>
<td>Amine</td>
<td>N6</td>
<td>Sigma</td>
<td>Sus.</td>
</tr>
<tr>
<td></td>
<td>Carboxylated</td>
<td>N7</td>
<td>Polysciences</td>
<td>Sus.</td>
</tr>
<tr>
<td>Aluminium oxide</td>
<td>7nm</td>
<td>N8</td>
<td>Krahn Chemie</td>
<td>Dry</td>
</tr>
<tr>
<td></td>
<td>50nm</td>
<td>N9</td>
<td>AHT*</td>
<td>Dry</td>
</tr>
<tr>
<td></td>
<td>300nm</td>
<td>N10</td>
<td>AHT*</td>
<td>Dry</td>
</tr>
<tr>
<td>Cerium oxide</td>
<td>N11</td>
<td>NAM**</td>
<td>Dry</td>
<td></td>
</tr>
<tr>
<td>Nickel oxide</td>
<td>N12</td>
<td>NAM**</td>
<td>Dry</td>
<td></td>
</tr>
<tr>
<td>Silicon oxide</td>
<td>N13</td>
<td>NAM**</td>
<td>Dry</td>
<td></td>
</tr>
<tr>
<td>Zinc oxide</td>
<td>N14</td>
<td>NAM**</td>
<td>Dry</td>
<td></td>
</tr>
<tr>
<td>Titanium dioxide</td>
<td>Rutile</td>
<td>N15</td>
<td>NAM**</td>
<td>Dry</td>
</tr>
<tr>
<td></td>
<td>Anatase</td>
<td>N16</td>
<td>NAM**</td>
<td>Dry</td>
</tr>
<tr>
<td>Silver</td>
<td>N17</td>
<td>NAM**</td>
<td>Dry</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N18</td>
<td>NAM**</td>
<td>Sus.</td>
<td></td>
</tr>
</tbody>
</table>

* AHT: Allied High Tech, ** NAM: Nanostructure and Amorphous Material Inc
**Characterisation of Physico-chemical Properties**

- **Particle shape** was analysed using LEO 1530 Scanning Electron Microscope (SEM) or Philips CM20 Transmission Electron Microscope (TEM) aspect ratio and mean size.

- **Surface area and porosity were measured using** TriStar 3000 BET. Thirteen measurements, including five surface areas based on different definitions, three pore volumes, three pore sizes as well as the mean size and particle density.

- **Particle size and size distribution** were analysed using a Malvern MasterSizer 2000. Size distribution, and seven other size properties (mass diameter, uniformity, specific surface area, surface area mean diameter and three mass diameters).

- **The free radical activities** were measured by EPR (Electron paramagnetic resonance) using the spin traps, DMPO and Tempone H separately. Tempone H has partial selectivity for trapping superoxide anions, whilst DMPO traps mainly hydroxyl radicals.

- **Particle reactivity in solution**, the dithiothreitol (DTT) consumption test as a reducing species was carried out. Since the DTT consumption test can only be conducted on dry powders, only fourteen of the panel of nanoparticles were assessed using this assay.

- **Metal Concentration** water soluble concentration of ten heavy metals (Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn and Cd) was measured.

- **Charge: z** potential was measured using Malvern Instrument’s Zetasizer Nano instrument.
### Total Number of Measurements

**13** • BET measurements (for 14 dry samples)

**2** • SEM/TEM measurements (for all samples)

**7** • laser diffraction size statistical measurements (Master Sizer 2000) for all samples

**84** • laser diffraction size distribution measurements

**2** • EPR measurements (for all samples)

**1** • reactivity measurement (for dry samples)

**10** • metal content measurements

**Total: 119 Measurements**

<table>
<thead>
<tr>
<th>Particle name</th>
<th>BET Measurements</th>
<th>M2000 Measurements</th>
<th>SEM Measurements</th>
<th>Zeta measurements</th>
<th>Other Measurements</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>A(1,1),A(1,2),…,A(1,12)</td>
<td>B(1,1),B(1,2)…</td>
<td>C(1,1),C(1,2)…</td>
<td>D(1,1),D(1,2)…</td>
<td>E(1,1),E(1,2)…</td>
</tr>
<tr>
<td>N2</td>
<td>A(2,1),A(2,2),…,A(2,12)</td>
<td>B(2,1),B(2,2)…</td>
<td>C(2,1),C(2,2)…</td>
<td>D(2,1),D(2,2)…</td>
<td>E(2,1),E(2,2)…</td>
</tr>
<tr>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
<td>…</td>
</tr>
<tr>
<td>N18</td>
<td>A(18,1),A(17,2),…</td>
<td>B(18,1),B(18,2)…</td>
<td>C(18,1),C(18,2)…</td>
<td>D(18,1),D(18,2)…</td>
<td>E(18,1),E(18,2)…</td>
</tr>
</tbody>
</table>
SEM and TEM images of the 18 nanoparticle samples
## The heavy metal concentration of each NP

<table>
<thead>
<tr>
<th>Particle names</th>
<th>Metal Concentration (µg/g)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ti</td>
</tr>
<tr>
<td>Carbon Black N1</td>
<td>0.00</td>
</tr>
<tr>
<td>Diesel Exhaust N2</td>
<td>2.320</td>
</tr>
<tr>
<td>Japanese Nanotubes N3</td>
<td>0.000</td>
</tr>
<tr>
<td>Fullerene N4</td>
<td>0.000</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N5</td>
<td>0.000</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N6</td>
<td>4.368</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N7</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N8</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N9</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N10</td>
<td>0.000</td>
</tr>
<tr>
<td>Cerium Oxide N11</td>
<td>0.000</td>
</tr>
<tr>
<td>Nickel Oxide N12</td>
<td>0.000</td>
</tr>
<tr>
<td>Silicon Oxide N13</td>
<td>0.000</td>
</tr>
<tr>
<td>Zinc Oxide N14</td>
<td>0.000</td>
</tr>
<tr>
<td>Titanium Dioxide Rutile N15</td>
<td>0.000</td>
</tr>
<tr>
<td>Titanium Dioxide Anatase N16</td>
<td>0.000</td>
</tr>
<tr>
<td>Silver N17</td>
<td>0.000</td>
</tr>
<tr>
<td>Silver N18</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Characterisation of Particle Size and Shape

LEO 1530 Gemini FEGSEM
Scanning Electron Microscopy

Mean size obtained by measuring the particles on the SEM/TEM image: \textit{Mean\_Size\_SEM}
Aspect ratio obtained from SEM/TEM images: \textit{Aspect\_Ratio}
**BET Measurements for 14 dry powder samples**

Miromeritics TriStar 3000

**Surface Area Technique:** Gas Adsorption

**Minimum Surface Area:** 0.01 m²/g

**Support Gases:** Nitrogen and Helium

**Measurement Parameters:**
- BET Measurement
- Adsorption Isotherms
- Desorption Isotherms
- t-plots
- Pore Volume Analysis
- Pore Size Analysis

<table>
<thead>
<tr>
<th>Surface Area (m²/g)</th>
<th>Pore Volume (cm³/g)</th>
<th>Pore Sizes (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Single point surface area at P/Po = 0.197 (m²/g) – SPSA</td>
<td>• Single point adsorption total pore volume of pores at P/Po = 0.98: SP_TP_VOP</td>
<td>• Adsorption average pore width: BET_PW</td>
</tr>
<tr>
<td>• BET Surface Area: BET_SA</td>
<td>• BJH Adsorption cumulative volume of pores between 17,000 Å and 3000.000 Å diameter: BJH_AC_VOP</td>
<td>• BJH Adsorption average pore diameter: BJH_A_PD</td>
</tr>
<tr>
<td>• Langmuir Surface Area: LM_SA</td>
<td>• BJH Desorption cumulative volume of pores between 17,000 Å and 3000.000 Å diameter: BJH_DC_VOP</td>
<td>• BJH Desorption average pore diameter: BJH_D_PD</td>
</tr>
<tr>
<td>• BJH Adsorption cumulative surface area of pores between 17,000 Å and 3000.000 Å diameter: BJH_AC_SAOP</td>
<td>• Mean size calculated from BET surface area: Mean_Size_BET</td>
<td>• Particle density: Density</td>
</tr>
<tr>
<td>• BJH Desorption cumulative surface area of pores between 17,000 Å and 3000.000 Å diameter: BJH_DC_SAOP</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**QualityNano Training School**

**14 dry powder samples, no available for three suspensions (N5, N6, and N7) and the new Silver sample.**
Malvern Mastersizer 2000
Type of Analyser: Laser Diffraction
Sample Volume: 1000ml, 150ml (Hydro G and S)
Measurement Range: 20nm – 2000μm

The information about mean Size and Size distribution, can be obtained from this sizer.
Characterisation of NPs-Mastersizer

Mastersizer 2000 Measurements: 91**

- $D[4,3]$ – the Volume Weighted Mean or Mass Moment Mean Diameter, represented by $D[4,3]$
- Uniformity: a measure of the absolute deviation from the median, represented by **Uniformity**
- Specific surface area: the total area of the particles divided by the total weight, **SSA_2000**
- $D[3,2]$ – the Surface Weighted Mean or Surface Area Moment Mean Diameter, represented by $D[3,2]$
- $D(v,0.5)$ – the size in microns at which 50% of the sample is smaller and 50% is larger. This value is also known as the Mass Median Diameter or the median of the volume distribution, represented by $D(0.5)$
- $D(v,0.1)$ – the size of particle below which 10% of the sample lies, represented by $D(0.1)$
- $D(v,0.9)$ – the size of particle below which 90% of the sample lies, represented by $D(0.9)$
- Size distribution: there are 100 values. Among them there are 16 values equal to zero for all particles. So 84 size distribution attributes are used for further analysis

** Note: available for all samples
Characterisation of NPs-Zeta Potential

Malvern ZetaSizerNano instrument

Size ranges: 0.6nm - 6µm

Zeta Potential

Suitable: suspensions
Toxicity Results

Cytotoxicity by LDH Release

Apoptosis

Viability

Necrosis

Pro-inflammation effects

Haemolysis
Principal Component Analysis of Cytotoxicity Data

- **LDH**
- **Apoptosis**
- **Viability**
- **Total-Toxicity**
- **Necrosis**
- **Haemolysis**

MTT

QualityNano Training School
Principal Component Analysis of Cytotoxicity Data

- The major cluster contains low toxicity particle samples
  - Aminated beads (N6), zinc oxide (N14), Japanese Nanotubes (N3) and nickel oxide (N12) are separate from the major cluster.
  - Aminated beads (N6) has the highest toxicity values in nearly all assay results (LDH, apoptosis, necrosis, haemolytic, MTT and cell morphology assays).
  - Zinc oxide (N14) has high toxicity value in LDH, apoptosis, necrosis and inflammation assays.
  - Japanese nanotubes (N3) have high toxicity values in viability and MTT assays.
  - Nickel oxide (N12) has shown highly toxic in LDH and haemolytic assays results..

The distinction of the four particles, N6, N14, N3 and N12, from other particles is clear.

Figure: Plot of the first and second PCs based on PCA analysis of toxicity data
The toxicity results of N6, N3, N12 and N14 which are likely to be toxic
Next Step: Principal Component Analysis of Descriptors

The focus of the (Q)SAR Analysis is to identify the possible structural and compositional properties which may contribute to the toxicity of zinc oxide (N14), polystyrene latex amine (N6), Japanese nanotubes (N3) and nickel oxide (N12).

The theory is that there should be structural or/and compositional distinctions for these nanoparticles from the rest of the non-toxic particles.

PCA and clustering were then applied to the physicochemical descriptors with the hypothesis that the toxic nanoparticles might also be discriminated from the non-toxic particles based on the analysis of the measured physicochemical descriptors.

If this can be proved, further analysis can be conducted to identify the key physicochemical descriptors that lead to the observed toxicity.
Principal Component Analysis of Descriptors

Principal component analysis of the structural and compositional descriptors for the panel of 18 nanoparticles (excluding BET and DTT measured descriptors since they are not available for the four wet samples).

- The scaling operation was performed, 3 PCs were then applied.
- PC1-PC2-PC3 three dimensional plot shows that the 18 nanoparticles were clearly grouped into clusters.
- The largest cluster contains particles that showed low toxicity values.
- Japanese Nanotubes (N3), nickel oxide (N12) and zinc oxide (N14) and diesel exhaust particles (N2) are outside this low toxicity sample cluster.
- There must be another reason that leads to high toxicity value for N6.

Figure: PCA analysis of the physicochemical descriptors for the panel of 18 nanoparticles (excluding BET and DTT measured descriptors since they are not available for the four wet samples).
To examine the clustering results more clearly, PC1-PC2, PC1-PC3 and PC2-PC3 two dimensional plots are also examined.

It is PC2 that separates N3 (Japanese nanotubes), N12 (nickel oxide) and N14 (zinc oxide) from the rest samples, as can be seen from PC1-PC2 and PC2-PC3 plots.
The contribution plot for, PC1, PC2 and PC3 is shown in the Figure, we are mainly interested in the contribution to PC2 (solid red colour), since it is PC2 that discriminated Japanese nanotubes (N3) nickel oxide (N12) and zinc oxide (N14) from the rest particle samples.

The largest contributions to PC2 are: aspect ratio measured by SEM imaging, and volume weighted mean, uniformity, and D(0.9), all measured by Mastersizer, as well as Ni and Zn metal contents.

Further analysis is required.
Results

• In order to determine exactly which physicochemical descriptors lead to high toxicity values of N3, N12, N14 and N6, several PC analysis were performed.

• For the 14 dry particles, PC analysis with BET and DTT measurements gave similar result with the first PC analysis in which BET and DTT data was excluded.

Figure: Principal component analysis of the structural and compositional descriptors (all such descriptors including BET and DTT measurements) for the 14 dry nanoparticles.
• To identify the relative contribution of each metal concentration to PC1 and PC2, the contribution plot to PC1 and PC2 is shown in the Figure.

• Ni content made the most important contribution to PC1, suggesting that nickel content may be the main reason for the high toxicity of N12 (nickel oxide).
• While Zn content made the most obvious contribution to PC2, suggesting that zinc content may be the reason for the high toxicity of N2 (diesel exhaust particles) and N14 (zinc oxide).
## The heavy metal concentration of each NP

<table>
<thead>
<tr>
<th>Particle names</th>
<th>Ti</th>
<th>V</th>
<th>Cr</th>
<th>Mn</th>
<th>Fe</th>
<th>Co</th>
<th>Ni</th>
<th>Cu</th>
<th>Zn</th>
<th>Cd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbon Black N1</td>
<td>0.00</td>
<td>0.057</td>
<td>0.00</td>
<td>0.305</td>
<td>0.00</td>
<td>0.002</td>
<td>0.263</td>
<td>1.161</td>
<td>9.687</td>
<td>0.008</td>
</tr>
<tr>
<td>Diesel Exhaust N2</td>
<td>2.320</td>
<td>0.312</td>
<td>16.47</td>
<td>20.81</td>
<td>208.4</td>
<td>0.508</td>
<td>4.940</td>
<td>2.235</td>
<td>3181</td>
<td>0.387</td>
</tr>
<tr>
<td>Japanese Nanotubes N3</td>
<td>0.00</td>
<td>0.100</td>
<td>0.00</td>
<td>0.161</td>
<td>13.10</td>
<td>0.000</td>
<td>0.299</td>
<td>0.123</td>
<td>0.460</td>
<td>0.001</td>
</tr>
<tr>
<td>Fullerene N4</td>
<td>0.00</td>
<td>0.084</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.534</td>
<td>0.142</td>
<td>4.837</td>
<td>0.000</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N5</td>
<td>0.00</td>
<td>1.172</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.625</td>
<td>19.56</td>
<td>0.000</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N6</td>
<td>4.368</td>
<td>0.779</td>
<td>0.343</td>
<td>0.199</td>
<td>0.000</td>
<td>0.006</td>
<td>0.440</td>
<td>18.23</td>
<td>58.71</td>
<td>0.010</td>
</tr>
<tr>
<td>Polystyrene Latex Beads N7</td>
<td>0.00</td>
<td>1.018</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>2.588</td>
<td>22.323</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N8</td>
<td>0.00</td>
<td>0.000</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.714</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N9</td>
<td>0.00</td>
<td>0.031</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.056</td>
<td>0.099</td>
<td>4.987</td>
<td>0.000</td>
</tr>
<tr>
<td>Aluminium Oxide N10</td>
<td>0.00</td>
<td>0.062</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.023</td>
<td>0.088</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Cerium Oxide N11</td>
<td>0.00</td>
<td>0.004</td>
<td>1.058</td>
<td>0.128</td>
<td>0.000</td>
<td>0.000</td>
<td>0.616</td>
<td>0.191</td>
<td>112.9</td>
<td>0.007</td>
</tr>
<tr>
<td>Nickel Oxide N12</td>
<td>0.00</td>
<td>0.056</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.042</td>
<td>16963</td>
<td>0.286</td>
<td>0.000</td>
<td>0.038</td>
</tr>
<tr>
<td>Silicon Oxide N13</td>
<td>0.00</td>
<td>0.088</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>14.163</td>
<td>0.106</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Zinc Oxide N14</td>
<td>0.00</td>
<td>0.096</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.336</td>
<td>0.047</td>
<td>8185</td>
<td>9.711</td>
</tr>
<tr>
<td>Titanium Dioxide Rutile N15</td>
<td>0.00</td>
<td>0.083</td>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.286</td>
<td>0.276</td>
<td>39.623</td>
<td>0.000</td>
</tr>
<tr>
<td>Titanium Dioxide Anatase N16</td>
<td>0.00</td>
<td>0.104</td>
<td>0.00</td>
<td>0.160</td>
<td>0.000</td>
<td>0.006</td>
<td>0.692</td>
<td>0.303</td>
<td>402.3</td>
<td>0.000</td>
</tr>
<tr>
<td>Silver N17</td>
<td>0.00</td>
<td>0.111</td>
<td>0.00</td>
<td>0.503</td>
<td>0.000</td>
<td>0.000</td>
<td>0.475</td>
<td>0.597</td>
<td>17.867</td>
<td>0.019</td>
</tr>
<tr>
<td>Silver N18</td>
<td>0.00</td>
<td>0.096</td>
<td>0.00</td>
<td>0.256</td>
<td>0.000</td>
<td>0.059</td>
<td>0.561</td>
<td>162.8</td>
<td>194.8</td>
<td>0.037</td>
</tr>
</tbody>
</table>
Conclusions

**Identifying the factor that discriminates N6 (polystyrene latex amine) from N5 (unmodified) and N7 (carboxylated)**

PCA analysis: PCA analysis of metal contents only, or PCA analysis of structural descriptors only, or PCA analysis of both structural and compositional descriptors together, could not discriminate N5, N6 and N7 from the largest clusters of non-toxic particle samples.

**PCA Conclusion: neither structural descriptors nor metal contents made N6 different fro N5 and N7 in toxicity**

It was at this stage of analysis that we suspected that there are other possible reasons for toxicity specific to N6: either it was due to the charge difference, or due to N6 being amine, or both.

*zeta potential* for all three polystyrene latex samples.

<table>
<thead>
<tr>
<th></th>
<th>N6 (amine)</th>
<th>N5 (unmodified)</th>
<th>N7 (carboxylated)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>zeta potential</strong></td>
<td>37.8, 37.5, and 40.3 (mV), positive</td>
<td>-36.2, -38.8 and -36.8, negative</td>
<td>-54.9, -55.3, and -58.6, negative</td>
</tr>
</tbody>
</table>

According to Verma et al. any particle with positive charge is likely to interact electrostatically since most cell surfaces and other biological membranes are negatively charged under physiological conditions(Verma et al. 2008).

**Therefore, it is most likely that the large positive charge of N6 contributed its observed high toxicity, despite it structurally similar to N5 and N7.**
Conclusions

• The SAR analysis generates the following conclusions:
  1. The most likely reason of high toxicity for Japanese nanotube (N3) is its shape (high aspect ratio).
  2. The most likely cause of high toxicity for zinc oxide (N14) and nickel oxide (N12) is their high contents of zinc and nickel.
  3. Although, Aminated beads (N6) has the highest toxicity values in nearly all assay results, it was always grouped closely with the other two polystyrene latex beads, unmodified beads (N5) and carboxylated beads (N7), which showed no toxicity. Therefore, the zeta potential of the three polystyrene latex was measured. The result suggests that the large positive charge of N6 contributed to its observed high toxicity.

• The work has shown that SAR based on molecular descriptors is a useful tool for describing factors influencing the toxicity of nanoparticles.

• A panel of eighteen nanoparticles is still considered to be too small, therefore, a QSAR model that can be applied to predict the toxicity of new nanoparticles cannot be developed with this small set of data.