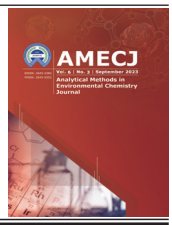
Anal. Methods Environ. Chem. J. 6 (3) (2023) 36-53

AMECJ

Research Article, Issue 3
Analytical Methods in Environmental Chemistry Journal
Journal home page: www.amecj.com/ir



Virtual chemical analysis and machine learning-based prediction of polyethylene terephthalate nanoplastics toxicity on aquatic organisms as influenced by particle size and properties

Christian Ebere Enyoh[a,†], Chidi Edbert Duru[b], Qingyue Wang[a], and Senlin Lu[c]

[a]*Graduate School of Science and Engineering, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 8570-338, Japan.*

[b]*Department of Chemistry, Faculty of Physical Sciences, Imo State University, PMB2000 Owerri, Nigeria.*

[C] *School of environmental and chemical engineering, Shanghai University, Shanghai 200444, China.*

ARTICLE INFO:

Received 27 May 2023

Revised form 30 Jul 2023

Accepted 22 Aug 2023

Available online 28 Sep 2023

Keywords:
Analytical methods,
Artificial neural networks,
Fish,
Health risks,
Plastic pollution,
Simulation,
Toxicity

ABSTRACT

This study focuses on the chemical analysis and prediction of Polyethylene Terephthalate (PET) nanoplastics toxicity on aquatic organisms, considering the influence of particle size and properties. The effect PET NPs of different sizes (1, 4, 9, 16 and 25 nm coded NP1 to NP5) on aquatic organisms such as *Terpedo californica* (electric ray fish) and *Danio rerio* (zebrafish) as model species was evaluated by virtual chemical techniques and machine learning methodology based on Multilayer Perceptrons Artificial Neural Networks (MLP ANN) and Support Vector Machine. The PET NPs was built and characterized *in silico* and then docked on the acetylcholinesterase (TcAChE) and cytochrome P450 (Zf CYP450) of the organisms, respectively. The results showed that the binding affinities of the NPs increased steadily from – 7.1 kcal mol^{-1} to – 9.9 kcal mol^{-1} for NP1 to NP4 and experienced a drop at NP5 (– 8.9 kcal mol^{-1}) for TcAChE. The Zf CYP450 also had a similar pattern ranging from -5.2 kcal mol^{-1} to -8.1 kcal mol^{-1}. The MLP ANN showed an accuracy of 85.9 % and 77.3 %. In comparison, SVM showed a better PET NPs toxicity prediction with an accuracy of 99.5 % and 99.4% based on the inherent properties of TcAChE and Zf CYP450, respectively.

1. Introduction

Plastic manufacturing has continuously expanded from 1.5 million metric tonnes in 1950 to 368 million tonnes in 2019 [1], and it is omnipresent in our daily lives. Most plastic manufactured worldwide is used for single-use items [2]. As a result, plastic wastes, mostly made of single-use plastics, are now a growing global problem for environmental degradation. Due to a dramatic increase in the use of single-use plastics like gloves and N95 or surgical masks, the new coronavirus epidemic has exacerbated plastic pollution [3]. By 2050, landfills and aquatic bodies will contain nearly 12,000 million metric tonnes of plastic trash, endangering marine and terrestrial ecosystems [1] if current trends in plastic manufacture and waste management are allowed to continue. Natural weathering processes can cause discarded plastic to break down into microplastics (MPs; 100 nm—5 mm size) and/or nanoplastics (NPs < 100 nm) [4]. NPs could be consumed by biota because of their small size [4]. NPs are a fast-evolving topic important in various sectors, including human toxicity, food and environmental study [4].

*Corresponding Author: Christian Ebere Enyoh
Email: cenyoh@gmail.com
https://doi.org/10.24200/amecj.v6.i03.249

Ingestion of NPs can expose aquatic biota, which can then accumulate the particles and experience adverse effects [5]. The harmful effects of NP exposure in aquatic species include embryotoxicity, hepatotoxicity, growth inhibition of microalgae and fish larvae, reduced shrimp and plankton life spans upon long-term exposure, deterioration of intestinal tissues in sea bass, and changes in feeding behaviour, metabolism, and innate immunity in fish by Banerjee et al., 2021. Human exposure to NPs happens when they consume tainted food or drink or breathe in airborne plastic particles [1,6]. It has not been possible to analyze the consequences of NPs exposure in humans directly [7]; however, MP has been found in human feces [8], placenta, in both the maternal and fetal regions, and amniotic membranes [9]. Studies have often translated toxicity in some aquatic organisms to humans because of similar genetic traits [10]. The aquatic organism Zebrafish, which shares genetic traits with mammals, has been utilized extensively in MPs research for assessing environmental toxicity and studying genetic developments [10, 11]. Zebrafish and humans have a lot in common physiologically and genetically. Zebrafish and humans have around 70 % of the same genes, and 84 % of the genes associated with human illnesses are also present in zebrafish [12]. There are reports that due to their great sensitivity to toxic contaminants, embryos and juveniles of Zebrafish can be used effectively to examine the detrimental impacts of environmental pollutants on aquatic creatures and the poisonous process [13]. The toxicity of chemicals to zebrafish mainly involved the cytochrome P450 (CYP) enzymes [14]. The cytochrome P450 (CYP) enzymes catalyze oxidative transformation, which results in the activation or inactivation of a wide range of endogenous and foreign substances. This has implications for both healthy physiology and disease processes. The oxidative biotransformation of these xenobiotics may decide those chemicals' cellular and organ targets [15]. Many compounds that cause developmental abnormalities, such as cardiovascular, neurological, and connective tissue disorders, are substrates for CYPs. The CYP enzymes involved in xenobiotic metabolism may also play a role in producing morphogenic molecules or maintaining regions free of them, thereby defining the temporal and spatial domains in which morphogens act and facilitating the region-specific adjustments required for successful development [15]. On the other hand, an electric rays' fish (ERF) was also considered a model species for aquatic organisms in this study to evaluate the toxicity of NPs to benthic organisms as they often exist in the coral reefs, mud, or sandy bottoms [16]. ERFs are renowned for having the ability to generate an electric discharge that may range in voltage from 8 to 220 volts, depending on the species, and are used to defend themselves as well as to shock prey [17]. The species of the genus Torpedo may be the most well-known [16]. Its name appears on the underwater torpedo weapon. The word "torpere" means "to stiffen or paralyse" in Latin, referring to the impact the fish has on those who contact it [18]. The ERF is often found in shallow water up to 100 metres on coral reefs, mud, or sandy bottoms [19]. It mostly consumes invertebrates and tiny reef fish [16].

To evaluate the toxicity of pollutants to this organism, acetylcholinesterase (AChE) is the primary target. The AChE is a primary cholinesterase in organisms that catalyzes the breakdown of acetylcholine and some other choline esters that function as neurotransmitters. It is found mainly at neuromuscular junctions, where it terminates synaptic transmission. Drugs or toxins that inhibit AChE led to the persistence of high concentrations of acetylcholine within synapses, leading to increased cholinergic signalling within the central nervous system, autonomic ganglia and neuromuscular junctions.

Compounds that inhibit AChE irreversibly may lead to convulsions, bronchial constriction, muscular paralysis, and death by asphyxiation [20]. One important factor that can influence the toxicity of NPs to aquatic organisms is their size. Studies have shown that size affects the rate of ingestion of NPs to shrimp [21] and toxicity to human gastric cells [22]. In this study, the toxicity

of polyethylene terephthalate (PET) NPs of different sizes to aquatic organisms was studied using a molecular docking (MD) approach coupled with machine learning (ML) methodology, using the CYP450 in zebrafish and AChE in ERF as primary targets. By simulating the interaction between a small molecule and a protein at the atomic level, the MD method enables us to describe how tiny molecules behave at the target proteins' binding site and better understand fundamental biological processes [23]. The two primary steps in the docking procedure are predicting the ligand structure, positioning and orientation inside these sites and assessing the binding affinity, which is translated into toxicity. On the other hand, a group of algorithms called "machine learning (ML)" are used to analyze data and learn without being specifically trained to use it for forecasting. ML contains a variety of algorithms, including supervised, semi-supervised, unsupervised, and reinforced, depending on the application situation [24]. Using the supervised algorithm, which recognizes data patterns, gains knowledge from observations, and generates predictions, the dataset has labelled inputs and outputs [25]. These are employed in problems of classification, regression, and forecasting variety [24]. Machine learning (ML) has expanded to include research into the toxicity of microplastics, predicting effects on cell, immune, and reproductive systems, and identifying the accumulation of materials in organisms based on material properties [26,27]. As a result, ML is expected to emerge as a valuable tool for quickly assessing the potential biological impact of NPs. In this study, the properties data set collected from NPs structures designed and analyzed *in silico* were labelled using supervised ML and used to predict the toxicity of the PET NPs to the studied aquatic organisms.

2. Material and Methods

2.1. Preparation of PET Nanoplastics (NPs)

The PET NPs of different sizes were prepared *in silico* using the 3D atomistic in Biovia Material Studio 8. Firstly, a single 3D isotactic ethylene terephthalate homopolymer molecular model was sketched. Depending on the size of NPs, the number of chain lengths and repeat units were determined. Four PET NPs of different sizes were prepared and compared with a single monomer unit of PET MPs. The structure, chemical formula and molecular weight are presented in Table 1.

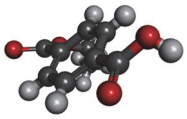
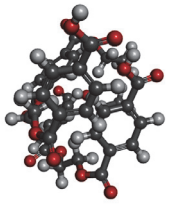
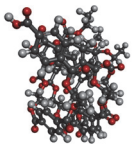
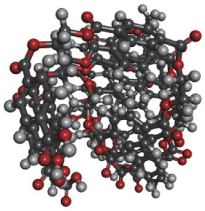
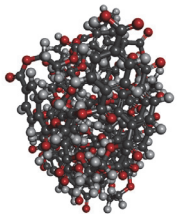
To increase the simulation's accuracy and efficiency, geometric optimization, namely energy reduction, is required for the developed PET NPs molecular models. The FORCITE module with COMPASS Forcefield was used to execute a geometry optimization work on the first formed molecular model with an unsatisfactory starting structure to reduce energy. The intelligent algorithm was optimized while the convergence tolerance energy and force were set to 0.001 kcal mol^{-1} and 0.5 kcal mol^{-1} A^{-1}, respectively, with 500 iterations. The optimized NPs structures were characterized to be sure they were in the nano-sized range. The results summary from the optimization and the characterization is presented in Table 3, which includes modifications to the total energy, surface area and occupied volume.

2.2. Effect of particle size on toxicity to aquatic organisms

2.2.1.Identification and preparation of enzyme targets

The 3D X-ray crystallographic structures of the acetylcholinesterase of *Terpedo californica* (electric ray fish) TcAChE with identity 1W76 and zebrafish cytochrome P450 (Zf CYP450) with identity 4R20 were retrieved from the protein data bank (PDB) which was shown in Figure1 [28]. The Chain A of the enzymes was used as a target to study the effect of chain length on nanoplastic toxicity. Their active sites were identified from the literature [14,29]. Removal of the interfering crystallographic water molecules and minimization of the protein was done using UCSF Chimera 1.14 [23,30,31].

Table 1. The PET NPs prepared in silico for the study

Samples	Chain length x repeat unit (nm)	Structure	Chemical formula	Molecular weight
PET 1	1 x 1 (1 nm)		$C_{10}H_{10}O_4$	194.18
PET NPs 2	2 x 2 (4 nm)		$C_{40}H_{36}O_{16}$	772.70
PET NPs 3	3 x 3 (9 nm)		$C_{90}H_{78}O_{36}$	1747.66
PET NPs 4	4 x 4 (16 nm)		$C_{160}H_{136}O_{64}$	3082.85
PET NPs 5	5 x 5 (25 nm)		$C_{250}H_{210}O_{100}$	4814.28

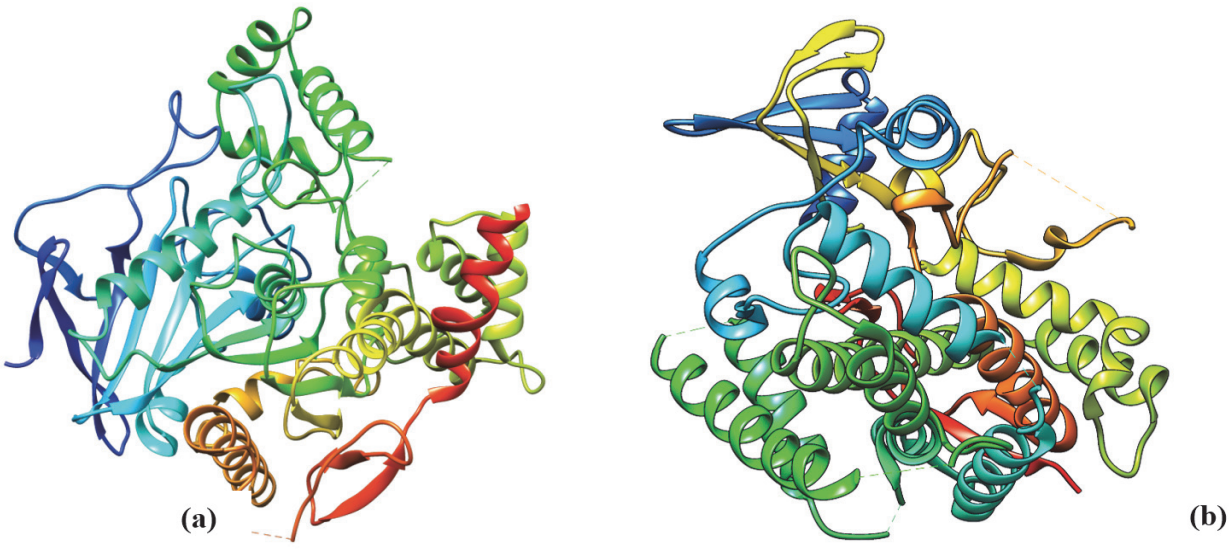


Fig. 1. Crystal structures of (a) TcAChE (b) Zf CYP450

2.2.2.Molecular docking studies

Site-directed docking of the nanoplastic compounds was performed on the active sites of the enzymes with Autodock Vina in PyRx software version 0.8 [32,33]. The amino acids at the active sites of the enzymes were identified in UCSF Chimera, and they were then selected and toggled on the enzyme surfaces in the Pyrx software. The specific sites on the receptors were set using the grid box with dimensions: center *x*: 88.622, center *y*: 55.838, center *z*: – 21.553, and size *x*: 30.496, size *y*: 27.705, size *z*: 18.659 for TcAChE, and center *x*: 3.640, center *y*: 5.221, center *z*: 50.783, and size *x*: 21.812, size *y*: 10.712, size *z*: 25.919 for Zf CYP450. At the end of the molecular docking, nine binding poses of the protein-ligand complex were generated, and their scoring results were also created. Hydrogen bonding and other hydrophobic interactions between the enzyme-ligand complex of the compounds were visualized using Biovia Discovery Studio 4.5 [34].

2.3. Machine Learning approach to model the toxicity of NPs to aquatic organisms

2.3.1.Artificial Neural Network (ANN)

This study uses Multilayer Perceptrons-based (MLP) Artificial Neural Networks (ANNs) as its machine learning strategy. ANNs are nonlinear models of interconnected "neurons," or units that can recognize patterns in various ways, including classification and prediction [25,26]. ANNs learn by seeing patterns in the data, and they store their newfound knowledge in weights, which are collections of connection strengths corresponding to regression coefficients. The MLP ANN employed in this work learns using backpropagation, where weights are adjusted following the processing of the whole data set or each datum. ANN weights evaluate the correlations between independent and dependent variables similarly to regression coefficients. However, ANN weights quantify local effects, whereas regression weights evaluate the global impacts of independent factors on dependent variables across all data [35].

An MLP comprises input, hidden, and output layers. Figure 2 presents an MLP ANN with four inputs (x1, x2, x3 and x4), one hidden-layer neuron and one output (f(x)). One can express the result of every hidden neuron (j) by utilizing the input values of Xi, weights, biases, and transfer function f(x) in the subsequent Equation 1.

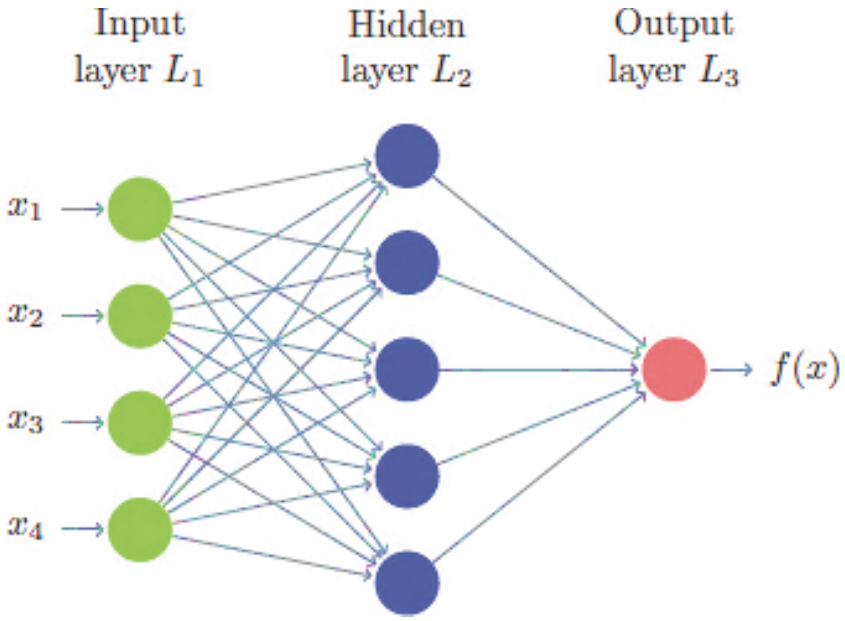


Fig. 2. An MLP ANN model with four input layers

Table 2. The range of data employed in ANN training

Parameter	Type	Minimum value	Maximum value
Binding affinity for AChE ($^{kcal\ mol-1}$)	Output	-7.1	-9.9
Binding affinity for CYP450 ($^{kcal\ mol-1}$)	Output	-5.2	-8.1
Molecular size (nm)	Input	1	25
Structure Energy ($^{kcal\ mol-1}$)	Input	13.51	200.16
Occupied volume ($Å^3$)	Input	168.32	3748.59
Surface area ($Å^2$)	Input	189.78	2148.33

$$H_j^o = f\left(\sum_{i=1}^{i} X_i \,.\, W_{ij} + b_j\right) \quad \text{(Eq. 1)}$$

The input and output variables used in this study for development are presented in Table 2. The input layer includes the independent variables, and the output layer contains the dependent variable. Equation 2 is used to express the final output (k) from the hidden layer result.

$$O_k^o = f\left(\sum_{i=1}^{i} H_j^o \,.\, W_{jk} + b_k\right) \quad \text{(Eq. 2)}$$

80 % of the data was used for training the ANN, while 20 % was used for testing. The batch training strategy, which produces the best results in data sets with minimal input variables, was used to train the network [36].

The MLP ANN makes an effort to calculate the weights connecting the input and output layers, representing the relative weights of the inputs into and outputs from the ANN "black box" or hidden layer [37]. In this study, the number of units in the hidden layer (excluding bias) varied for AChE and CYP450, respectively. At the same time, the activation function was based on hyperbolic tangent and linear. The arguments are transformed into values between -1 and +1 via the hyperbolic tangent, frequently employed in ANN modelling. This activation function is beneficial when the training data are normalized, as in the current research [35]. The hidden layer changes the input data into a collection of values and is then used by the output layer. The mathematical operations taking place in the hidden layer are what produce the ANN's nonlinear behaviour, making it a crucial part of the system [35,37]. In this study's case, the dependent variable, or binary classifications, is contained in the output layer. During training,, the ANN will examine the output layer's dependent variable and the input layer's independent variables to see how they specifically relate to one another. The hidden layer's nodes include mathematical functions that define the relationships. Once the connections (mathematical functions) have been established, the testing data will be used to validate them [38].

2.3.2.Support vector machine

SVM has several advantages compared to ANN, including its ability to avoid being trapped in local minima by mapping the nonlinear relationship between inputs and output(s), solving problems using only support vectors, and handling small data sets [37]. According to Yettou [39], the performance of an SVM model depends on the choice of kernel function and its parameters. The predicted output can be obtained using the SVM model as Equation 3.

$$y(x)_{pre} = \sum_{i=1}^{n} a_i \,.\, K\left(x_i, x_j\right) + b \quad \text{(Eq. 3)}$$

K(xi, xj) can be linear, polynomial, Gaussian or radial basis function kernel. αi and b denote the Lagrange multiplier and threshold parameter, respectively. In this study, the radial basis function kernel was used (Equation 4). The dataset used for the SVM is shown in Table 2. 20 support vectors were used in the model with a cost of 1 and gamma of 0.25. The gamma hyperparameter was selected before training the model and influenced the decision boundary's curvature.

$$f(x) = \sum_i^N a_i y_j \, exp\left(\frac{-||x-x_i||^2}{2\sigma^2}\right) + b \quad \text{(Eq. 4)}$$

2.3.3. Model evaluation metrics

Different statistical models, such as the average correlation factor (R^2), root mean square error (RMSE), and mean absolute error (MAE). The sum of square error (SSE) specified in Equations 5-8 was used to check and evaluate how well an ML model predicts the output (binding affinity, kcal mol^{-1}) [37].

$$R^2 = 1 - \frac{\sum(BA_{est} - BA_{exp})^2}{\sum(BA_{est} - \overline{BA}_{exp})^2} \quad \text{(Eq. 5)}$$

$$RMSE = \sqrt{\frac{1}{N}\sum(BA_{est} - BA_{exp})^2} \quad \text{(Eq. 6)}$$

$$MAE = \frac{1}{N}\sum(BA_{est} - BA_{exp}) \quad \text{(Eq. 7)}$$

$$SSE = \sum(BA_{est} - BA_{exp})^2 \quad \text{(Eq.8)}$$

Where:

- $(BA)_{est}$: is the estimated value of the binding affinity in kcal mol^{-1} by ANN model;
- $(BA)_{exp}$, is the experimental value of the binding affinity in kcal mol^{-1}; and
- : is the average value of binding affinity in kcal mol^{-1}

The goal should be to achieve the lowest error with (RMSE, MAE, and SSE) and the greatest with (R^2) correlations to obtain the optimal ANN model [37, 38].

3. Results and discussion

Although there is ample evidence that nanoplastics (NPs) are toxic to aquatic life, little is known about how these particles' sizes affect their toxicities. The binding affinity of NP2, NP3, and NP4 was higher than the control for TcAChE, indicating that they have a higher inhibitory potential at this site and, therefore, could impact more toxicity on the enzyme than the native inhibitor (-)-galanthamine. However, the abiraterone control for Zf CYP450 showed the highest toxicity with a binding affinity of -9.5 kcal mol^{-1} compared with the PET NPs. The validation of the model was tested by error analysis models, which confirmed that the SVM was at high accuracy. The study demonstrated that the size of PET NPs can influence their toxicity to aquatic organisms, and the surface area and reactivity (energy) of the NPs are important for their toxicity. Integrating chemistry analysis, machine learning, and environmental chemistry allows for accurately predicting Polyethylene Terephthalate (PET) nanoplastics toxicity on aquatic organisms, considering various particle sizes and properties.

3.1. Optimization and characterization of PET NPs

The prepared PET NPs were optimized and characterized. Geometry minimization is finding an arrangement in the space of a collection of atoms where the net inter-atomic force on each atom is acceptably close to zero and the position on the potential energy surface (PES) is a stationary point. Since this is the most stable configuration for the molecule and is most likely to occur in nature, geometry optimization aims to find the location where the energy is lowest [40]. The optimized energies, surface area and occupied volume were determined *in silico;* the results are presented in Table 3. After optimization, the PET NPs initial structure energy was reduced to 13.51, 91.78, 96.24, 200.16 and 176.47 kcal mol^{-1}, respectively, for 1 (1 nm) to 5 (25 nm). The most stable (geometry-optimized) forms of the NPs were used in the toxicity evaluation of aquatic organisms. The cornolly surface area of the nanoplastics is also presented in Figure 3. The surface area of the PET NPs ranged from 189.78 to 2148.33 $Å^2$, and the occupied volume of the PETs increased from

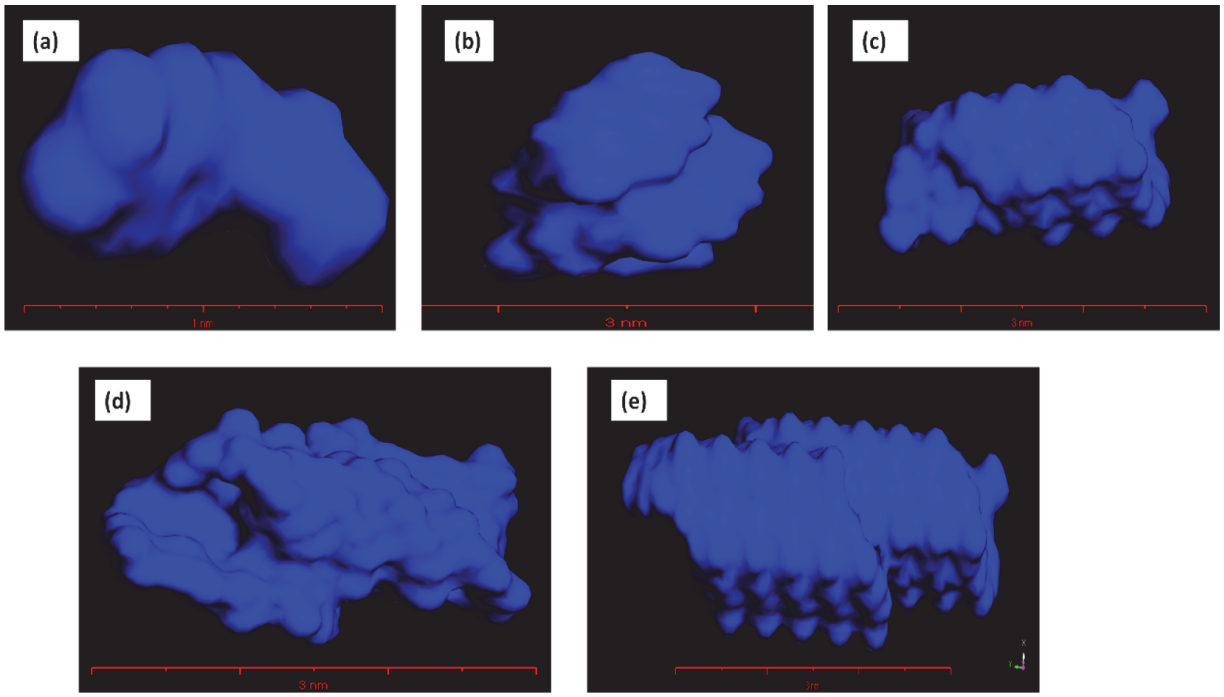


Fig. 3. Connolly surface plot of PET NPs for surface area and occupied volume determination. Figure (a) to (e) represents PET NPs 1 – 5.

Table 3. The PET NPs optimization and surface characterization results

Samples	Molecular Size (nm)	Energy (kcal mol-1) Initial	Final	Occupied volume (Å^3)	Surface area (Å^2)
PET NP 1	1	160.57	13.51	168.32	189.78
PET NPs 2	4	86874.58	91.78	734.18	651.47
PET NPs 3	9	137955.82	96.24	1441.39	1025.52
PET NPs 4	16	317658.71	200.16	2986.98	1783.66
PET NPs 5	25	8907.77	176.47	3748.59	2148.33

168.32 to 3748.59 Å^3 (Table 3). The surface area and occupied volume were determined based on the Connolly model [41R: SM]. Expectedly, the occupied volume and surface area increased with the PET NPs molecular sizes (Table 3). As the molecular sizes increase, the carbon to hydrogen increases as well, which has been shown to correlate to surface area [42R,43R; SM]. Enyoh [43] reported that the aging of PET microplastics increased the carbon-to-hydrogen ratio and, thus, the surface area of the material. Its surface area indicates how much overall space an object's surface takes up. Chemical reactions often proceed more quickly when a substance's surface area increases. However, the occupied volume measures the quantity of room inside an object. The surface area of the PET NPs confirmed that the PET NPs were nano-sized [44R; SM].

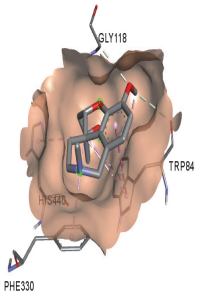
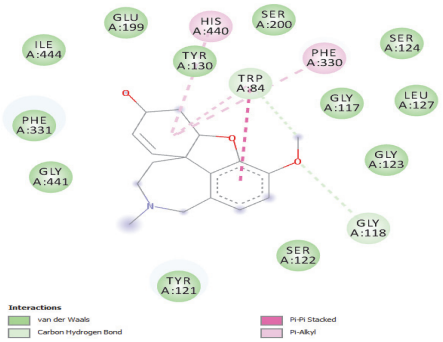
3.2. Effect of PET NPs size on toxicity to aquatic organisms through molecular docking

To evaluate the toxicity of NPs on aquatic organisms, the AChE of Electric ray fish (Terpedo californica) and CYP450 of Zebrafish (Danio rerio) were considered models. Acetylcholinesterase (AChE) is a primary cholinesterase in organisms that

catalyzes the breakdown of acetylcholine and some other choline esters that function as neurotransmitters. It is found mainly at neuromuscular junctions, where it terminates synaptic transmission. Drugs or toxins that inhibit AChE led to the persistence of high concentrations of acetylcholine within synapses, leading to increased cholinergic signalling within the central nervous system, autonomic ganglia and neuromuscular junctions. Compounds that inhibit AChE irreversibly may lead to convulsions, bronchial constriction, muscular paralysis, and death by asphyxiation [20]. However, CYP450 of zebrafish has increasingly been used in drug discovery and toxicology screening [30]. The CYP enzymes generally catalyze the oxidative transformation of many endogenous and exogenous chemicals, thereby functioning as a metabolizer of potentially toxic compounds in organisms. The CYP genes in zebra fish have been shown to have a high degree of sequence similarity with those of humans [15].

The binding affinities and the interactions of the enzyme-NP complex at the AChE active site are shown in Table 4. The decrease in binding affinity implies greater stability of the NP at the active site, which results in the inhibition of the enzyme activities. The binding affinities of the NPs increased steadily from – 7.1 kcal mol^{-1} to – 9.9 kcal mol^{-1} for NP1 to NP4 and experienced a drop at NP5 (– 8.9 kcal mol^{-1}). The binding affinity of NP2, NP3, and NP4 was higher than the control, indicating that they have a higher inhibitory potential at this site and, therefore, could impact more toxicity on the enzyme than the native inhibitor (-)-galanthamine. All the polymer chains from NP2 to NP4 fitted well into the binding pocket of AChE. Including an additional monomer unit in NP5 resulted in a constraint which manifested as a drop in the binding affinity of this polymer unit. A similar observation was reported in an experimental study by [22]. They evaluated polystyrene NPs of particle sizes (50, 100, 200, 500, 1000 or 5000 nm). They reported a toxicity trend in gastric cells in order of 50 > 5000 > 1000 > 500 > 200 > 100 nm. [21] reported an increase in mortality (toxicity to adult daggerblade grass shrimp *Palaemonetes pugio*) as MPs sizes increase. However, [45R; SM] found no effect of 200–600 µm PET MPs sizes on toxicity to *Scenedesmus* sp. The binding affinities of the different NPs and the control compound abiraterone at the Zf CYP450 active site are shown in Table 5. The control showed the highest toxicity with the binding affinity of -9.5 kcal mol^{-1} compared with the PET NPs, which ranged from -5.2 kcal mol^{-1} to -8.1 kcal mol^{-1}. The binding pocket of this protein could not accommodate two PET units from NP5, which manifested as a sharp drop in the binding affinity of the enzyme-NP5 complex. However, the toxicity pattern of the PET NPs on the Zf CYP450 active site is like AChE of Electric ray fish, indicating that increasing NP size will increase toxicity to a certain extent, followed by a reduction in toxicity due to inefficient enzyme-NP binding.

Table 4. The binding affinity of nanoplastic compounds on the TcAChE active site

PET NPs	NP-Protein complex	Binding affinity (kcal mol^{-1})	Protein-ligand interactions
Control		– 9.2	

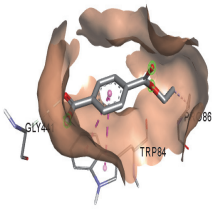
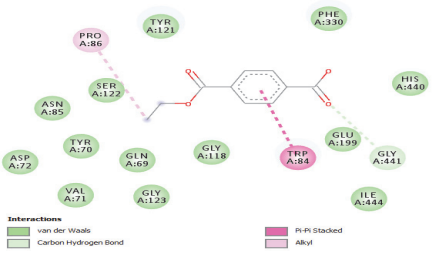
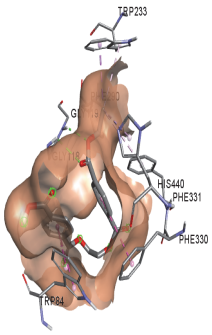
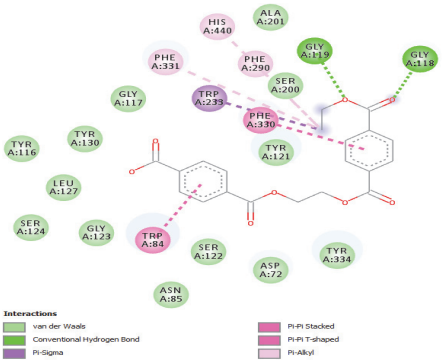
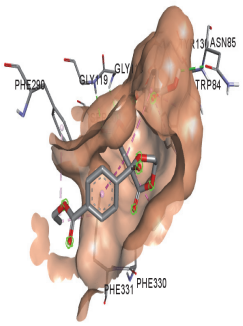
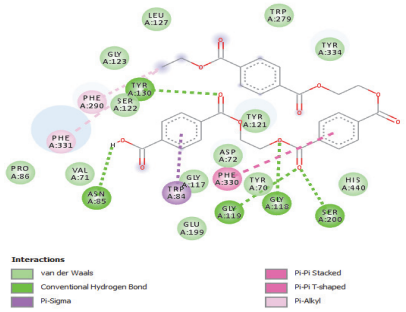
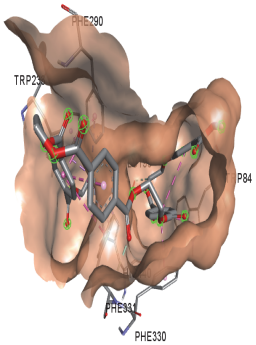
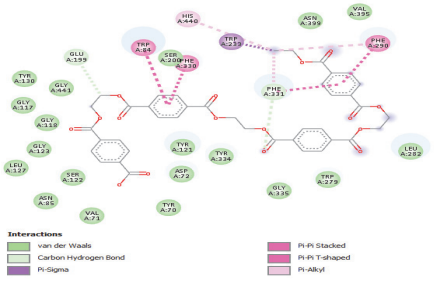
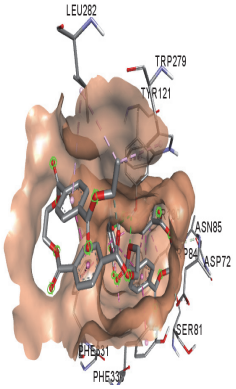
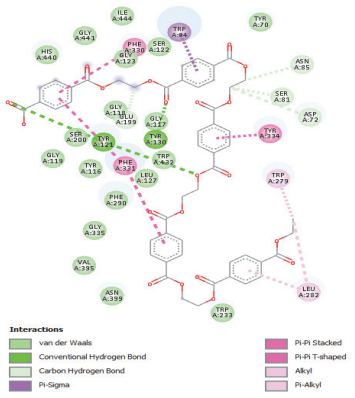
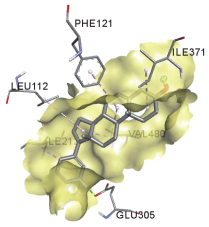
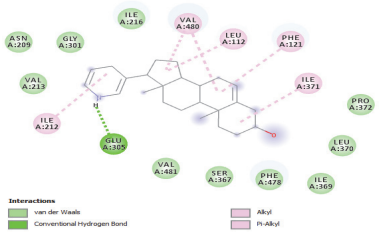
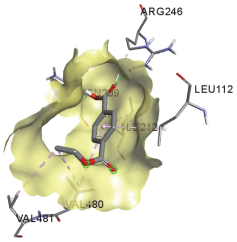
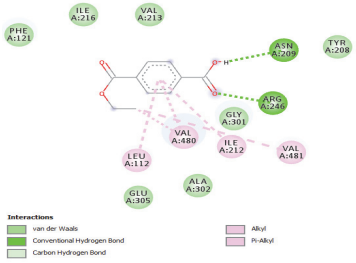
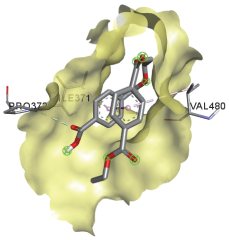
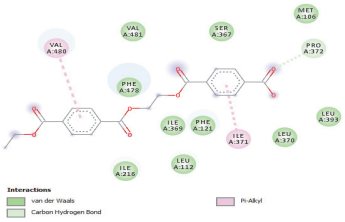
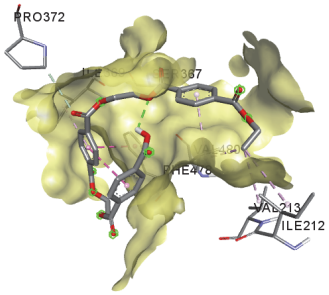
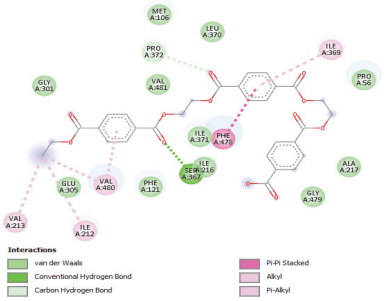
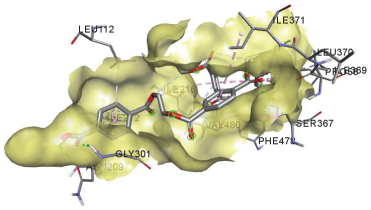
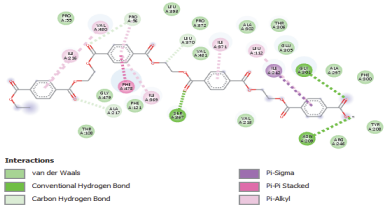
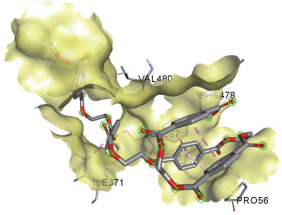
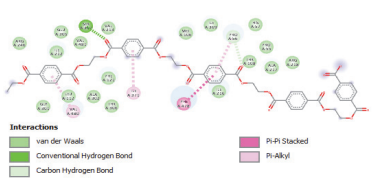
1		– 7.1	
2		– 9.3	
3		– 9.5	
4		– 9.9	
5		– 8.9	

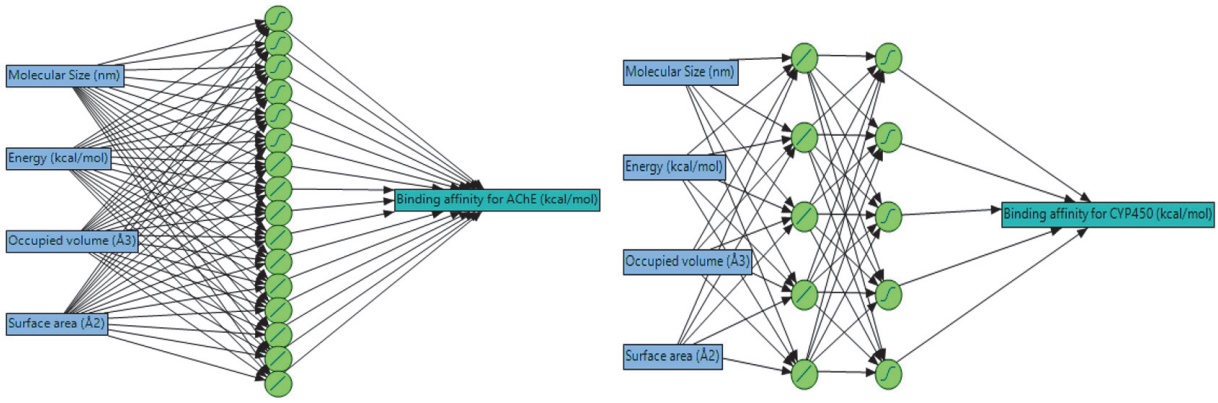
Table 5. The binding affinity of nanoplastics compounds on the Zf CYP450 active site

PET NPs	NP-Protein complex	Binding affinity (kcal mol^{-1})	Protein-ligand interactions
Control		– 9.5	
1		– 5.2	
2		– 5.8	
3		– 7.3	
4		– 8.1	
5		– 5.7	

3.3. Toxicity evaluation via machine learning methodology

The toxicity of NPs is based on their structural characteristics, such as their molecular Size (MS), optimized energy, occupied volume (OC) and surface area (SA) at the active sites of TcAChE and Zf CYP450. The results showing the artificial neural networks, identification of the most important property and linear correlation of predicted binding affinity (toxicity) from machine learning and actual values from molecular docking are presented in Figures 4 to 5. The ANN for the two enzymes differed (Figure 4); AChE had 15 hidden layers (excluding bias), while CYP450 had just five each in the first and second hidden layers. Classification boundaries are simpler to define in the hidden layer than in the original space, which may be considered a mapping to a higher dimensional space. Out of the 20 SVM models tested model number 17 and 4 were best to predict the binding affinity to Tc AChE (Table S1, moved to supplementary material, SM) and Zf CYP450 (Table S2, moved to supplementary material, SM) based on the smallest root average squared error (RASE) values. These models had a cost of 2.49, and gamma and support vector (SV) of 0.499 and 4, respectively for Tc AChE, while for Zf CYP450, cost, gamma, and support vector (SV) were 3.71, 0.19 and 4, respectively.

The correlation of the predictive toxicity value from ANN and actual values obtained from the molecular docking were significant, with coefficients of 0.859 and 0.773 for Tc AChE and Zf CYP450, respectively (Figure 5a and 5b). However, the SVM algorithm showed a better correlation with the actual with coefficients of 0.995 and 0.994 for Tc AChE and Zf CYP450, respectively (Figure 5c and 5d). Correlation coefficient values > 0.9 are considered significant and acceptable [46R,47R,48R; SM]. Figures 5e and 5f compare the experimental values and those predicted by a Support Vector Machine (SVM) model. The results demonstrate that the points predicted by the SVM model align perfectly with the trend of the experimental points, indicating that the SVM model can accurately model the non-linear behavior of the reactivity descriptor of the PET NPs. In contrast, the performance of an Artificial Neural Network (ANN) model in this regard is not as good as that of the SVM model. The comparison between the experimental and predicted values is an essential step in evaluating the performance of machine learning models. The accuracy of the model's predictions is critical in determining its effectiveness in modelling the underlying system. In this case, the SVM model is shown to outperform the ANN model in modelling the complex and non-linear behavior of the reactivity descriptor of the PET NPs. This significant result highlights the importance of selecting the appropriate machine-learning model for a specific application.



(a) **(b)**

Fig. 4. The artificial neural networks for predicting NPs toxicity based on their properties to (a) AChE and (b) CYP450.

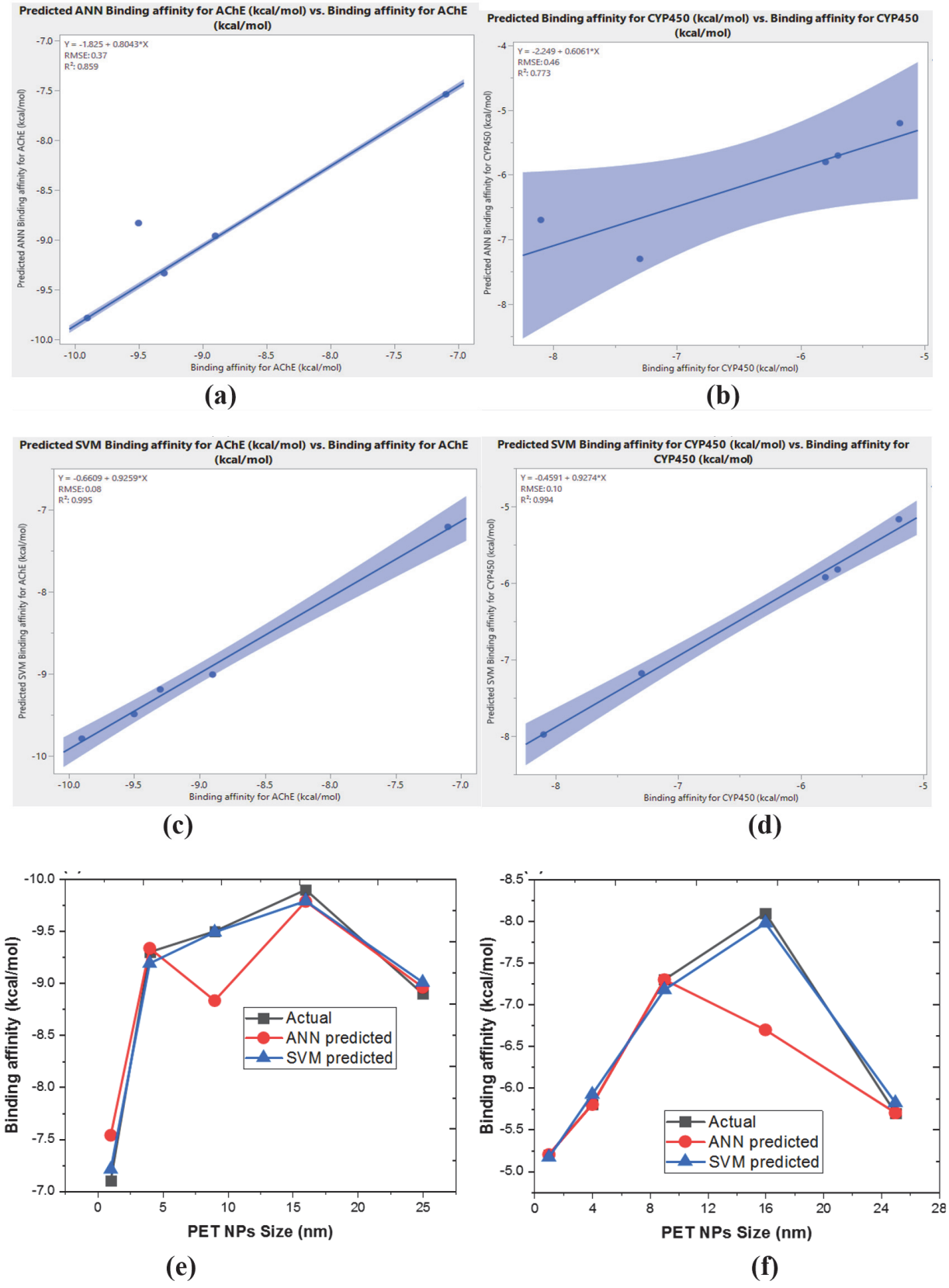


Fig. 5. The correlation of predicted binding affinity (toxicity) from ANN (a)-(b) and SVM (c)-(d) and actual values from molecular docking studies to AChE and CYP450. The comparative plot of the predicted binding affinity from ANN and SVM with actual values to (e) AChE and (f) CYP450

Furthermore, the SVM model's ability to accurately model the behavior of the reactivity descriptor of the PET NPs has significant implications for the field of nanotechnology. The reactivity descriptor is a critical property of the PET NPs that determines their reactivity and, therefore, their potential toxicity to aquatic organisms. Accurately modelling this property is essential for designing and optimizing these nanoparticles for various applications.

The most important NPs property for the toxicity varied with the proteins (Figures 6a and 6b). The NPs optimized energy and surface area were the main properties with normalized importance of 100 % that affect the toxicity of NPs at TcAChE and Zf CYP450 active sites, respectively. The NPs energy is indicative of their reactivity. Xie [49R; SM] recently found that nanoparticle chemical reactivity and surface area were important parameters influencing the toxicity of nanoparticles to algae. The results of a previous study using both in vivo and in vitro methodologies showed that the amount of surface area in contact with the mouse's biological system determines the response's amplitude, suggesting that surface chemistry phenomena are involved in biological reactivity [50R; SM]. Further, [51R; SM] reported that the surface area is the physiologically most efficient dosage meter for acute nanoparticle toxicity in the lung.

The efficacy of the constructed ANN and SVM model was additionally evaluated for reliability using error analysis models such as the root mean square error (RMSE), mean absolute error (MAE), and sum of square error (SSE). The error analysis results of the ANN and SVM are summarized in Table 6. The RMSE measures the difference between the toxicity values predicted by the model and the actual values [52R,53R; SM]. The values were 0.37 and 0.46 for AChE and CYP450, respectively, from the ANN, while 0.08 and 0.10 for SVM, indicating that the SVM is a better model. The MAE of each individual prediction error on all occurrences in the test set is the MAE of a model for the test set. The discrepancy between a forecast's expected value and its real value is known as a prediction error [54R; SM]. The ANN MAE for AChE and CYP450 were higher than those of SVM. Similar results were also obtained for the SSE. SSE is an accuracy metric that adds squared errors. When the data points are of equal magnitude, it is utilized to assess the forecasting model's accuracy [53R; SM]. The RMSE, MAE, and SSE values were lower for SVM, and the R^2 values were high for TcAChE and Zf CYP450, respectively, compared to ANN. These indicate that the system's prediction is more reliable for the SVM than the ANN model [55R; SM].

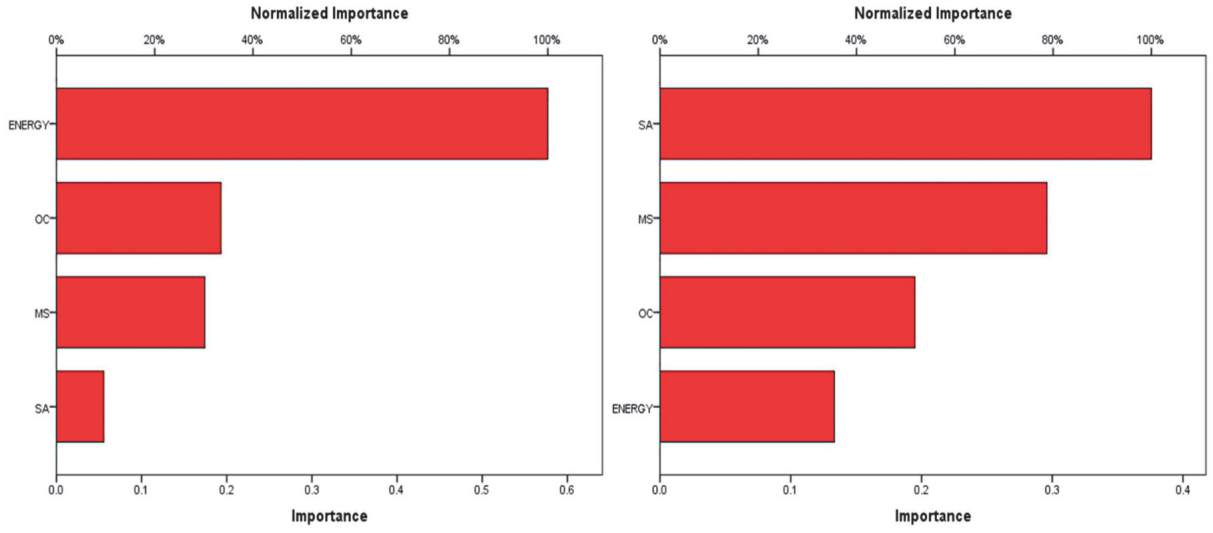


Fig. 6. The most important of NPs property from the ANN in predicting toxicity to (a) AChE (b) CYP450

Table 6. Error analysis results for the ANN

ML Algorithm	Model	AChE	CYP450
ANN	R^2	0.859	0.773
	RMSE	0.37	0.46
	MAE	0.55	0.65
	SSE	0.53	0.64
SVM	R^2	0.995	0.995
	RMSE	0.08	0.10
	MAE	0.06	0.07
	SSE	0.02	0.03

4. Conclusion

The present study examined the toxicity of Polyethylene Terephthalate Nanoplastics (PET NPs) to aquatic organisms through molecular docking and machine learning. The toxicity of the PET NPs increased with increasing size to a certain limit (16 nm), after which there was a drop in its toxicity due to inefficient enzyme-NP binding. Using the characteristics data from the PET NPs, the factors responsible for its toxicity were evaluated by the Machine Learning approach based on the Artificial Neural Network (ANN) and Support Vector Machine (SVM) model. The model was built for each component to predict the toxicity behavior of the PET NPs to zebrafish and electric ray fish, respectively. The predicted data obtained using SVM was at high accuracy for support vectors of 4, with $R^2 > 0.99$ and showed that surface area and reactivity (energy) were the most important properties for PET NPs toxicity. The study confirmed that the size of PET NPs can influence their toxicity to aquatic organisms, and the surface area and reactivity (energy) of the NPs are important for their toxicity.

5. Supplementary Materia l (SM)

Table S1 and S2 presented in SM. Also, more references from 41R- 55R moved to SM.

6. Conflicts of Interest

There are no conflicts of interest to declare.

7. Funding

This study was partially supported by the Special Funds for Basic Research (B) (No.22H03747, FY2022-FY2024) of Grant-in-Aid for Scientific Research of the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT).

8. References

[1] A.W. Verla, C.E. Enyoh, E.N. Verla, K. Nwanorh. Microplastic–toxic chemical interaction: a review study on quantified levels, mechanism and implication, SN Appl. Sci. 1 (2019a) 1400. https://doi.org/10.1007/s42452-019-1352-0

[2] Y. Chen, A.K. Awasthi, F. Wei, Q. Tan, J. Li, Single-use plastics: Production, usage, disposal, and adverse impacts, Sci. Total Environ., 752 (2021) 141772. https://doi.org/10.1016/j.scitotenv.2020.141772

[3] S.A.L. Patrício, J.C. Prata, T.R. Walker, A.C. Duarte, W. Ouyang, D. Barcelò, Increased plastic pollution due to COVID-19 pandemic: Challenges and recommendations, Chem. Eng. J., 405 (2021) 126683. https://doi.org/10.1016/j.cej.2020.126683

[4] C.E. Enyoh, Q. Wang, T. Chowdhury, W. Wang, S. Lu, K. Xiao, M.A.H Chowdhury, New analytical approaches for effective quantification and identification of nanoplastics in environmental samples, Processes, 9 (2021) 2086. https://doi.org/10.3390/pr9112086

[5] C. E. Enyoh, L. Shafea, A. W. Verla, E. N. Verla, W. Qingyue, T. Chowdhury, M. Paredes, Microplastics exposure routes and toxicity studies to ecosystems: An overview,

Environ. Anal. Health Toxicol., 35(1) (2020) 1–10. https://doi.org/10.5620/eaht.e2020004.

[6] C.E. Enyoh, A.W. Verla, E.N. Verla, Airborne microplastics: a review study on method for analysis, occurrence, movement and risks, Environ. Monit. Assess., 191 (2019) 668. https://doi.org/10.1007/s10661-019-7842-0

[7] C.E. Enyoh, W. Qingyue, V.A. Wirnkor, T. Chowdhury, Index models for ecological and health risks assessment of environmental micro-and nano-sized plastics, AIMS Environ. Sci., 9 (2022) 51-65. https://doi.org/10.3934/environsci.2022004

[8] P. Schwabl, S. Köppel, P. Königshofer, T. Bucsics, M. Trauner, T. Reiberger, B. Liebmann, Detection of various microplastics in human stool: A prospective case series, Ann. Inter. Med., 171 (2019) 453–457. https://doi.org/10.7326/M19-0618

[9] A. Ragusa, A. Svelato, C. Santacroce, P. Catalano, V. Notarstefano, O. Carnevali, F. Papa, M. C. A. Rongioletti, F. Baiocco, S. Draghi, E. D'Amore, D. Rinaldo, M. Matta, E. Giorgini, Plasticenta: First evidence of microplastics in human placenta, Environ. Int., 146 (2021) 106274. https://doi.org/10.1016/j.envint.2020.106274

[10] A.P. Araújo, T. Marinho, T. Lopes, Toxicity evaluation of the combination of emerging pollutants with polyethylene microplastics in zebrafish: Perspective study of genotoxicity, mutagenicity, and redox unbalance, J. Hazard. Mater., (2022) 432. https://doi.org/10.1016/j.jhazmat.2022.128691

[11] J. Bhagat, L. Zang, N. Nishimura, Y. Shimada, Zebrafish: An emerging model to study microplastic and nanoplastic toxicity, The Sci. total environ., 728, (2020) 138707. https://doi.org/10.1016/j.scitotenv.2020.138707

[12] R.L. Bailone, H.C.S. Fukushima, V. Fernandes, Zebrafish as an alternative animal model in human and animal vaccination research, Lab. Anim. Res., 36 (2020) 13. https://doi.org/10.1186/s42826-020-00042-4

[13] M. Ignacio, K. Le Menach, M. Devier, M.P. Cajaraville, H. Budzinski, A. Orbea, Screening of the toxicity of polystyrene nano- and microplastics alone and in combination with benzo(a)pyrene in brine Shrimp Larvae and Zebrafish embryos, Nanomater., 12 (2022) 941. https://doi.org/10.3390/nano12060941

[14] P.S. Pallan, L.D. Nagy, L. Lei, E. Gonzalez, Structural and kinetic basis of steroid 17α,20-lyase activity in teleost fish cytochrome P450 17A1 and its absence in cytochrome P450 17A2, J. Biol. Chem., 290 (2015) 3248-268. https://doi.org/10.1074/jbc.M114.627265

[15] J.V. Goldstone, A.G. McArthur, A. Kubota, Identification and developmental expression of the full complement of Cytochrome P450 genes in Zebrafish, BMC Genom., 11 (2010) 643. https://doi.org/10.1186/1471-2164-11-643

[16] P.R. Last, W.T. White, M.R. de Carvalho, B. Séret, M.F.W. Stehmann, G.J.P. Naylor, Rays of the world. CSIRO Publishing, Comstock Publishing Associates, i-ix, pp. 1-790, 2016. https://www.cornellpress.cornell.edu/book/9781501705328/rays-of-the-world/#bookTabs=1

[17] S.W. Michael, Reef sharks and rays of the world. A guide to their identification, behavior, and ecology, Sea Challengers Monterey California publisher, 107 pages, 1993. https://doi.org/10.1017/S0025315400034998

[18] M. R. Aidan, ReefQuest Centre for Shark Research, Electric Rays publisher, 2008. http://www.elasmo-research.org/education/shark_profiles/torpediniformes.htm

[19] C. Sommer, W. Schneider, J.M. Poutiers, FAO species identification field guide for fishery purposes, the living marine resources of Somalia, FAO publisher, Rome, 376 pages, 1996. https://www.fao.org/3/y0770e/y0770e.pdf

[20] M.B. Colović, D.Z. Krstić, T.D. Lazarević-Pašti, A.M. Bondžić, V.M. Vasić, Acetylcholinesterase inhibitors: pharmacology and toxicology, Curr. Neuropharmacol., 11 (2013) 315-335. https://

doi.org/10.2174/1570159X11311030006

[21] A. D. Gray, J. E. Weinstein, Size- and shape-dependent effects of microplastic particles on adult daggerblade grass shrimp (*Palaemonetes pugio*), Environ. Toxicol. Chem., 36 (2017) 3074–3080. https://doi.org/10.1002/etc.3881

[22] A. Banerjee, L.O. Billey, W.L. Shelver, Uptake and toxicity of polystyrene micro/nanoplastics in gastric cells: Effects of particle size and surface functionalization, PLOS ONE, 16 (12) (2021) e0260803. https://doi.org/10.1371/journal.pone.0260803

[23] C.E. Duru, I.A. Duru, C.E. Enyoh, In silico binding affinity analysis of microplastic compounds on PET hydrolase enzyme target of Ideonella sakaiensis, Bull. Natl. Res. Cent., 45 (2021) 104. https://doi.org/10.1186/s42269-021-00563-5

[24] V. Zhou, Machine learning for beginners: An introduction to neural networks, Medium, 2019. https://towardsdatascience.com/machine-learning-for-beginners-an-introduction-to-neural-networks-d49f22d238f9

[25] C. Duru, C. Enyoh, I.A. Duru, M.C. Enedoh, Degradation of PET nanoplastic oligomers at the novel PHL7 target: Insights from molecular docking and machine learning, J. Niger. Soc. Phys. Sci., 5 (2023) 1154–1154. https://doi.org/10.46481/JNSPS.2023.1154

[26] F. Yu, X. Hu, Machine learning may accelerate the recognition and control of microplastic pollution: Future prospects, J. Hazard. Mater., 432 (2022) 128730. https://doi.org/10.1016/j.jhazmat.2022.128730

[27] X. Wu, Z. Qixing, M. Li, H. Xiangang, Machine learning in the identification, prediction and exploration of environmental toxicology: challenges and perspectives, J. Hazard. Mater., 438 (2022) 129487. https://doi.org/10.1016/j.jhazmat.2022.129487.

[28] C. E. Enyoh, Q. Wang, P. E. Ovuoraye, T. O. Maduka, Toxicity evaluation of microplastics to aquatic organisms through molecular simulations and fractional factorial designs, Chemosphere, 308(Pt 2) (2022) 136342. https://doi.org/10.1016/j.chemosphere.2022.136342

[29] H.M. Greenblatt, C. Guillou, D. Guénard, A. Argaman, S. Botti, B. Badet, The complex of a bivalent derivative of galanthamine with torpedo acetylcholinesterase displays drastic deformation of the active-site gorge: implications for structure-based drug design, J. Am. Chem. Soc., 126 (2004) 15405-15411. https://doi.org/10.1021/ja0466154

[30] R.T. Peterson, C.A. Macrae, Systematic approaches to toxicology in the zebrafish, Annu. Rev. Pharm. Toxicol., 52 (2012) 433-453. https://doi.org/10.1146/annurev-pharmtox-010611-134751

[31] C.E. Duru, Duru I.A., A. Bilar, Computational investigation of sugar fermentation inhibition by bergenin at the pyruvate decarboxylate isoenzyme 1 target of *Scharomyces cervisiae*, J. Med. Plants Stud., 8(6) (2020) 21-25. https://doi.org/10.22271/plants.2020.v8.i6a.1225

[32] C.E. Duru, Duru I.A., A.E. Adegboyega, In Silico identification of compounds from *Nigella sativa* seed oil as potential inhibitors of SARS-CoV-2 targets, Bull. Natl. Res. Cent., 45 (2021) 57. https://doi.org/10.1186/s42269-021-00517-x

[33] C.E. Enyoh, O.M. Tochukwu, C. E. Duru, S.C. Osigwe, C.B.C. Ikpa, Q. Wang, In silico binding affinity studies of microbial enzymatic degradation of plastics, J. Hazard. Mater. Adv., 6 (2022)100076. https://doi.org/10.1016/j.hazadv.2022.100076.

[34] BIOVIA, Dassault Systemes, San Diego, Discovery studio modeling environment, 2020. https://docslib.org/doc/9937570/biovia-discovery-studio%C2%AE-2020

[35] C.E. Enyoh, C.E. Duru, E. Prosper, Q. Wang, Evaluation of nanoplastics toxicity to the human placenta in systems, J. Hazard. Mater., 446 (2023) 130600. https://doi.org/10.1016/j.jhazmat.2022.130600

[36] H. Tang, K. C. Tan, Z. Yi, Neural networks: Computational models and applications. Heidelberg, Germany: Springer, 2007. https://link.springer.com/book/10.1007/978-3-540-69226-3

[37] C.E. Enyoh, Q. Wang, L. Senlin, Optimizing the efficient removal of ciprofloxacin from aqueous solutions by polyethylene terephthalate microplastics using multivariate statistical approach, Chem. Eng. Sci., 278 (2023) 118917. https://doi.org/10.1016/j.ces.2023.118917

[38] C.E. Enyoh, P. Ovuoraye, O. Isiuku, C. Igwegbe, Artificial neural network and response surface design for modeling the competitive biosorption of pentachlorophenol and 2,4,6-trichlorophenol to canna indica L. in Aquaponia, Anal. Methods in Environ. Chem. J., 6 (2023) 79-99. https://doi.org/10.24200/amecj.v6.i01.228

[39] A. Yettou, M. Laidi, A. El Bey, S. Hanini, M. Hentabli, O. Khaldi, and M. Abderrahim Ternary Multicomponent Adsorption Modelling Using ANN, LS-SVR, and SVR Approach – Case Study, Kem. Ind., 70 (2021) 509−518. https://doi.org/10.15255/KUI.2020.071 KUI-36/2021

[40] C.E. Enyoh, Q. Wang, W. Weiqian, C. Tanzin, H.R. Mominul, I. Md. Rezwanul, Sorption of per- and polyfluoroalkyl substances (PFAS) using Polyethylene (PE) microplastics as adsorbent: Grand canonical Monte Carlo and molecular dynamics (GCMC-MD) studies, Int. J. Environ. Anal. Chem., (2022) 1-19. https://doi.org/10.1080/03067319.2022.2070016

(References from 41R-55R showed in SM)