

Experimental–computational study of carbon nanotube effects on mitochondrial respiration: in silico nano-QSPR machine learning models based on new Raman spectra transform with Markov–Shannon entropy invariants

Michael González-Durruthy^{†‡}, Luciane C. Alberici[§], Carlos Curti[§], Zeki Naal^{||}, David T. Atique-Sawazaki^{||}, José M. Vázquez-Naya[⊥], Humberto González-Díaz^{#○}, and Cristian R. Munteanu^{*⊥Δ}

[†] Institute of Biological Science (ICB) and [‡] Post-Graduate Program in Physiological Sciences, Universidade Federal do Rio Grande–FURG, 96270-900, Rio Grande, RS, Brazil

[§] Department of Physical Chemistry, Faculty of Pharmacy of Ribeirão Preto, and ^{||} Department of Chemistry, Faculty of Philosophy, Sciences and Letters, University of São Paulo (USP), 14040-903 Ribeirão Preto, SP, Brazil

[⊥] RNASA-IMEDIR, Computer Science Faculty, University of A Coruña, Campus de Elviña s/n, 15071 A Coruña, Spain

[#] Department of Organic Chemistry II, Faculty of Science and Technology, University of the Basque Country UPV/EHU, 48940, Leioa, Bizkaia, Spain

[○] IKERBASQUE, Basque Foundation for Science, 48011, Bilbao, Bizkaia, Spain

^Δ Instituto de Investigación Biomédica de A Coruña (INIBIC), Complejo Hospitalario Universitario de A Coruña (CHUAC), A Coruña, 15006, Spain

Abstract

The study of selective toxicity of carbon nanotubes (CNTs) on mitochondria (CNT-mitotoxicity) is of major interest for future biomedical applications. In the current work, the mitochondrial oxygen consumption (E3) is measured under three experimental conditions by exposure to pristine and oxidized CNTs (hydroxylated and carboxylated). Respiratory functional assays showed that the information on the CNT Raman spectroscopy could be useful to predict structural parameters of mitotoxicity induced by CNTs. The in vitro functional assays show that the mitochondrial oxidative phosphorylation by ATP-synthase (or state V3 of respiration) was not perturbed in isolated rat-liver mitochondria. For the first time a star graph (SG) transform of the CNT Raman spectra is proposed in order to obtain the raw information for a nano-QSPR model. Box–Jenkins and perturbation theory operators are used for the SG Shannon entropies. A modified RRegrs methodology is employed to test four regression methods such as multiple linear regression (LM), partial least squares regression (PLS), neural networks regression (NN), and random forest (RF). RF provides the best models to predict the mitochondrial oxygen consumption in the presence of specific CNTs with R^2 of 0.998–0.999 and RMSE of 0.0068–0.0133 (training and test subsets). This work is aimed at demonstrating that the SG transform of Raman spectra is useful to encode CNT information, similarly to the SG transform of the blood proteome spectra in cancer or electroencephalograms in epilepsy and also as a prospective chemoinformatics tool for nanorisk assessment.



INTRODUCTION

The boom in research on carbon nanotube (CNT) shaped nanoparticles has led to the development of more powerful designs and synthesis methods.¹ Currently, several scientific reports highlight a major impact of adverse/toxic effects induced by carbon nanomaterials on critical subcellular components, mainly mitochondria, which are responsible for the maintenance of the bioenergetic balance of ADP/ATP, redox, and cellular homeostasis in all eukaryotic organisms.^{2,3}

In this regard, carbon nanotubes have attracted attention for their high ability to accumulate in the mitochondrial matrix from several tissues and cells based on a peculiar mitotropic behavior. The important role of mitochondria to regulate intracellular ROS-levels based on the complete reduction of molecular oxygen by the respiratory complexes, has been extensively characterized and associated with several chronic pathological processes, such as neurodegenerative diseases (Alzheimer, Parkinson, Epilepsy), cardiovascular conditions, and cancer, which currently have high levels of morbidity and mortality, and where mitochondrial dysfunction mechanisms have been directly involved.^{2,3}

Mitochondrial events such as the dissipation of the membrane potential, the generation of reactive oxygen species and the release of caspase-activating proteins are closely linked to the mechanisms of cell death by apoptosis and necrosis. Lipophilic compounds such as carbon nanomaterials (single- and multiwalled CNTs (SWCNTs, MWCNTs)), with high lipid–water partition coefficients and enough access to the mitochondrial membranes could induce cell death mediated by mitochondrial mechanisms. The presence of carbon nanotubes near the respiratory chain can disrupt the normal flow of electrons along the respiratory complexes by decreasing the proton gradient and ATP synthesis. It is well-known that the mitochondria are essential elements in controlling the death or survival of the cell and they are, therefore, an important pharmacological and toxicological target that can be considered in the planning and evaluation of new carbon nanomaterials as potentially cytoprotective or cytotoxic based on their mitochondrial effects.^{2,3} Consequently, a certain concern was raised about the toxicity/safety rates of these new materials with emphasis on the respiratory system.⁴

In this context, our hypothesis is that new structural information (CNT nanodescriptors) obtained from Raman spectra, based on the criteria of rational drug design and mitochondrial medicine, could improve efficient therapeutic strategies against the aforementioned diseases, considering the chemical/pharmacological modulation of mitochondrial respiratory mechanisms.

Following this idea, it is well-known that Raman spectroscopy is one of the most important experimental techniques for the characterization, detection, biological interactions, and/or toxicity of CNTs. This is due to the fact that the D band feature in the CNT Raman spectra, with a peak at approximately 1350 cm^{-1} , is commonly associated with the presence of topological defects and type of functionalization (chemical oxidation) in the carbon lattice, which is known to be able to influence the interaction properties between the CNTs and the protein complexes of the mitochondrial electron transport chain, depending on the physical–chemical characteristics of carbon nanotubes.⁵

On the other hand, polarographic mechanistic assays provide a quick and reproducible means of measuring the rate of oxygen consumption by mitochondria isolated from different tissues.⁶ However, there are no precedents of this methodology applied to the evaluation of potential toxicity of CNTs. Some in vitro studies have demonstrated that CNTs exert cytotoxicity after their accumulation in the mitochondrial matrix and/or by affecting the function of mitochondrial proteins of the inner membrane.⁷ Previous in vitro research on drugs and environmental pollutants using submitochondrial particles (respiratory chain complexes I, II, III, IV; ADP/ATP translocator, ATP synthase/ATPase) to predict the toxic impact of 92 different xenobiotics showed a strong correlation with toxicity in humans and pointed out that the mitochondrial area was a relevant model for studying the relative toxicity of many xenobiotics.⁸ Recent efforts have been made to create a unified ontology for the annotation of data about nanomaterial safety entitled eNanoMapper.⁹

The combination of different methods proved to be a powerful tool in designing nanoscale systems such as CNTs, iron nanoparticles or micelle nanoparticles.¹⁰ Thus, quantitative structure–property/activity relationship (QSPR/QSAR) methods have been used as complementary tools for experiments by providing theoretical nano-QSPR prediction models.^{11–28} Since not all similar molecules have similar properties, the purpose is to define the *small* structural changes of molecules. In order to describe these differences, a PT-QSPR method has been proposed:²⁹ it combines QSPR/QSAR with perturbation theory (PT). Thus, the PT-QSPR models have been applied to complex molecular systems using a variation of multiple experimental boundary conditions: chemical reactivity, drug metabolism, vaccine peptide epitopes, metabolic networks, micelle nanoparticles,^{29–32} cytotoxicity of nanoparticles,³³ ecotoxicity and cytotoxicity of uncoated and coated nanoparticles,^{28,29} and antibacterial profiles of nanoparticles.³⁴ The classic QSPR/QSAR approach could be used with the Raman spectra descriptors and the observable output, but it limits the model information to the CNT Raman spectral data, without taking into the account the experimental conditions and time. Thus, the current study includes the information referring to the moving averages of the descriptors in specific experimental conditions, such as perturbation to the expected values of the observable output.

The design and development of novel carbon nanomaterials are currently expensive and complex processes. Thus, the new quantitative structure–activity relationship paradigm (or QSAR/QSPR tools) has become an important methodology for the prioritization/optimization of nanomaterials, as an alternative with less impact on health and environment in the nanoscience context (nanotoxicology).³⁵ Herein, a crucial step in QSPR is to express structural properties in a quantitative way, which is not always straightforward.^{35,36} Therefore, QSPR model can be seen as a mathematical function that predicts the structure of a single or complex system using physicochemical parameters which numerically describe its essential properties. To this end, Gonzalez-Díaz et al.²⁹ proposed a general purpose PT-QSPR method combining a QSPR chemoinformatics approach and PT. PT-QSPR approaches are very useful for the study of complex molecular biosystems with simultaneous multiple experimental boundary conditions.

González-Díaz et al.²⁹ applied the PT-QSPR analysis linked to chemical reactivity studies, drug metabolism (ADME–pharmacokinetic parameters), immunotoxicity tests, metabolic networks, metal, and CNT nanoparticles. Toropova et al.³⁷ published a nano-QSPR model on pristine MWCNTs to study the genotoxicity under multiple experimental conditions. Recently, González-Durruthy et al.,³⁸ using a PT-nano-QSPR approach, have been able to predict the mitochondrial swelling inhibition (mitochondrial permeability transition pore mitoprotective activity) induced by oxidized CNT in multiple experimental conditions. The authors concluded that oxidized CNT could modulate the mitochondrial ROS-production involved in mitochondrial dysfunction.³⁸

The main assumption of QSAR/QSPR approaches in general is that similar molecules have similar properties. Consequently, small structural changes (“perturbations”) should correlate linearly with small changes on the values of their properties (biological activities). The QSPR perturbation model is able to find out the exact solution of a problem (physicochemical and/or biological property) and continues adding small corrections to predict a solution for a related problem, without knowing an exact solution.^{39,40} In this context, currently there are no precedents for the application of this chemoinformatics methodology combining experimental and biochemical assays in isolated rat-liver mitochondria, including mechanistic explanations, to predict the potential effects of CNT on mitochondrial respiration toward biomedical applications based on Raman spectroscopy structural information.

The current study proposes the use of CNT Raman spectra as the main molecular information. Raman spectroscopy is a technique that provides the chemical fingerprint of molecules as vibrational, rotational, and other low-frequency modes. Thus, it is used for the study of biomolecular systems and nanoscale structure such as DNA,⁴¹ proteins,⁴² antibodies,⁴³ and carbon nanotubes (CNTs).⁴⁴ The Raman spectra can be transformed into Shannon entropies of the star graph, similar to a Fourier spectra transform. The use of star graphs as a graphical method to encode molecular/signal information has been proven for protein prediction⁴⁵⁻⁴⁷ and nucleic acid⁴⁸ function, as well as cancer prediction using blood protein mass spectra⁴⁹ and epileptic seizure prediction using electroencephalogram (EEG) signals.⁵⁰ The end point (mitochondrial respiration or E3) used in the QSPR/QSAR analyses is a predictive function of the mitochondrial oxygen consumption ($E3_{pred}$), which is the mitochondrial respiration under carbon nanotubes exposure obtained from the experimental data set as a function of Raman spectra transformed into star graph Shannon entropies indices (Sh_k) of CNT individual members of the tested CNT family. The expected values of the mitochondrial oxygen consumption ($E3_{exp}$) are the moving averages of the Raman spectra transformed into Sh_k indices under different experimental conditions and times of exposure. Points have been used as features (CNT nanodescriptors) to find the best PT-QSPR regression model that can predict the carbon nanotube effects on mitochondrial oxygen consumption in the presence of specific CNTs.

MATERIALS AND METHODS

General Workflow

The main aims of this paper are (a) development of a new transform of Raman spectra into Shannon entropy information indices (Sh_k), (b) measurement of the Raman spectra of a set of CNTs, (c) measurement of biological activity of these CNTs, and (d) use of the Sh_k values to predict the biological activity (mitochondria oxygen consumption or E3). In order to accomplish these objectives, a general workflow was proposed. The following list describes the steps of this workflow (Figure 1):

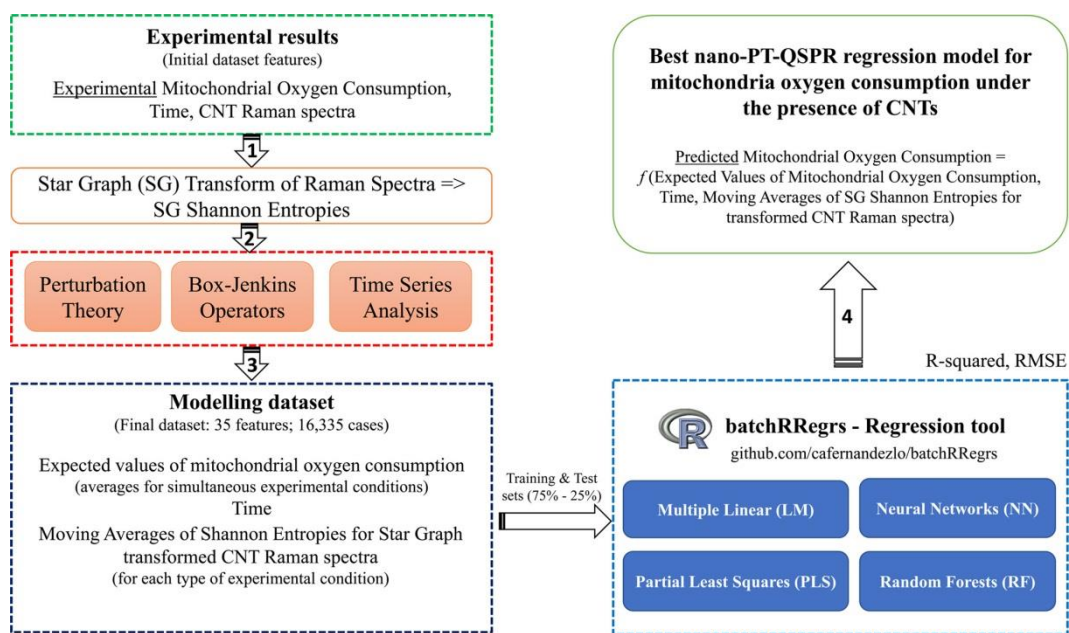


Figure 1. Methodology flowchart for nano-PT-QSPR models for mitochondrial oxygen consumption in the presence of CNTs.

- (1) Experimental measurements of the mitochondrial oxygen consumption under different experimental conditions with different types of CNTs
- (2) Transformation of CNT Raman spectra into star graph Shannon entropies
- (3) Use of nano-PT-QSPR methodology to calculate the expected values of the mitochondrial oxygen consumption and the moving averages of the Shannon entropies under different experimental conditions
- (4) Search of regression nano-PT-QSPR models using batchRRRegrs (an R tool for regressions)

Even if the RAMAN spectra need to be measured or predicted so that new CNTs could be used in the model, there is no need to measure the biological activity (E3) of CNTs in mitochondria, which is a difficult assay. In addition, it should be pointed out that the CNTs from the current study are commercial samples of CNTs. It means that the CNT samples do not contain single-molecule structures. On the contrary, each sample has many different molecules of CNTs with the same function type (H, OH, COOH) and the same percentage of this function, but in many different positions.

Therefore, the calculation of merely theoretical descriptors is not applicable for the current study because it does not involve a unique CNT molecule, but a mixture of many CNT molecules with similar, but different structures. In this sense, the Raman spectra are a good solution because they capture all the variations in the structural patterns of the CNTs into a sample.

EXPERIMENTAL SECTION

Raman Spectra

Raman spectra were measured using a Renishaw Micro-Raman Spectroscopy System (Renishaw plc, Wotton-under-Edge, UK) at room temperature at a laser excitation wavelength of 514 nm (2.33 eV). All reactions were quenched to room temperature before Raman spectra were recorded in order to identify the characteristic peaks in the position of 1580 cm^{-1} (G band of graphite) and the peak in the 1350 cm^{-1} (D band of defects) approximately associated with the presence of disorder and/or vacancy defects in the CNT structure produced by chemical oxidation in the graphite structure (oxidized CNT with OH and COOH functional groups), as shown in Figure 2.

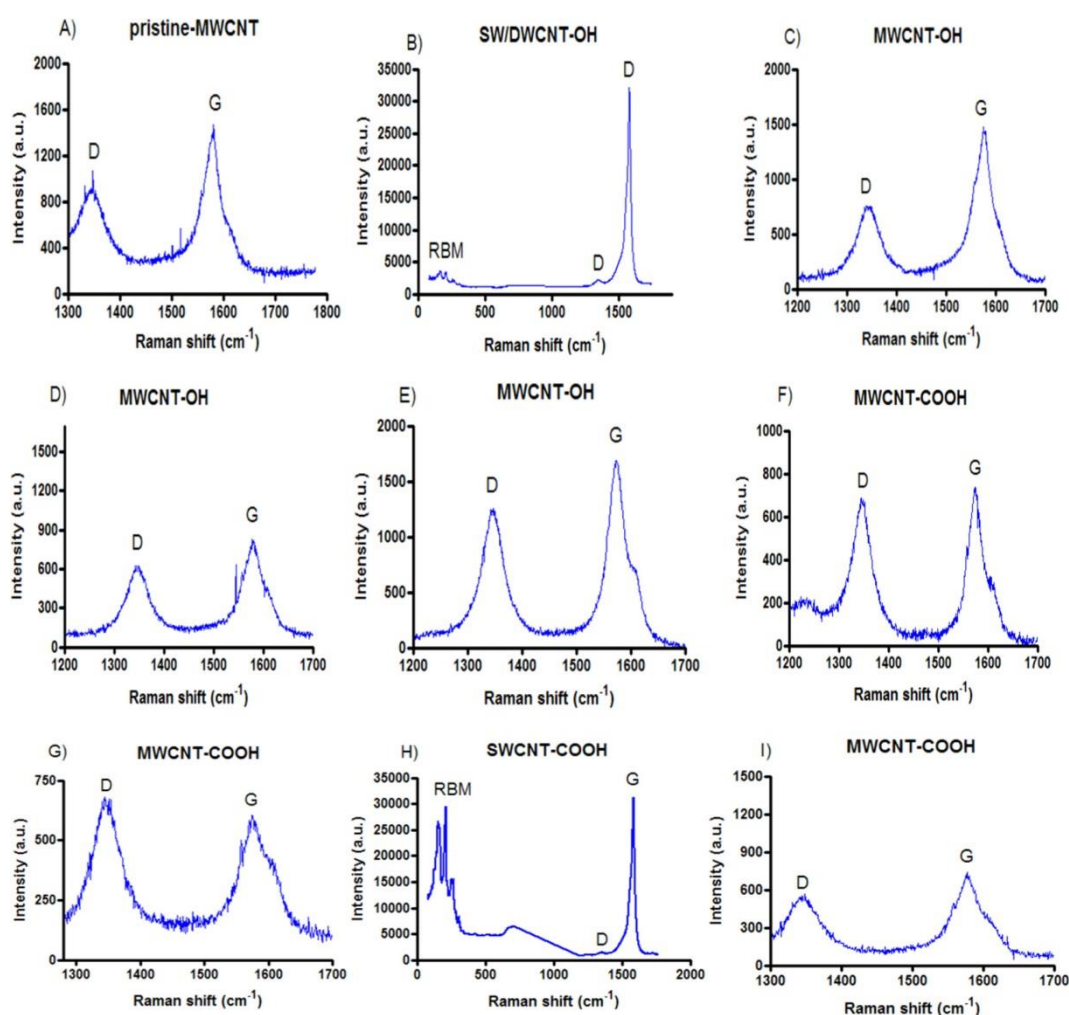


Figure 2. Raman spectra of carbon nanotubes used in this study. (A) Pristine MWCNT (CNT1), (B) SW/DWCNT-OH (CNT2), (C) MWCNT-OH (CNT3), (D) MWCNT-OH (CNT4), (E) MWCNT-OH (CNT5), (F) MWCNT-COOH (CNT6), (G) MWCNT-COOH (CNT7), (H) SWCNT-COOH (CNT8), (I) MWCNT-COOH (CNT9) (see Materials and Methods section).

Reagents and Solutions

Sucrose, ethylene glycol-bis(β -aminoethyl)-*N,N,N',N'*-tetraacetic acid (EGTA), KCL, potassium succinate (plus 2 μ M rotenone), K_2HPO_4 , piperazine-*N'*-2-ethanesulfonic acid (hepes-KOH). All the other reagents were commercial products of the highest purity grade available. For mitochondrial functional respiratory assays, pristine and functionalized carbon nanotubes (MWCNT, [SWCNT + DWCNT]-OH, MWCNT-OH, MWCNT-COOH, SWCNT-COOH) were dissolved in dimethyl sulfoxide (DMSO) and Milli-Q water in individual stock solutions, prepared at 1 mg/mL. The CNT family was provided by Cheaptubes Company (<http://cheaptubes.com/shortohcnts.htm>; see Table 1).

Table 1. Physical-Chemical Parameters of the CNT Family

<i>n</i>	CNT properties ^a		<i>W_i</i> (%)		<i>D_i</i> (nm)		<i>L_i</i> (μ m)	<i>P_i</i> (%)	<i>C_i</i> (S/cm)
	type	function	min	max	min	max			
1	MWCNT				8	8	0.5–2	>95	<1.5
2	mixed SW/DWCNT	OH	0	3.96	1	4	0.5–2	>95	<1.5
3	MWCNT	OH	0	3.86	1	8	0.5–2	>95	<1.5
4	MWCNT	OH	3	4	10	20	0.5–2	>95	<1.5
5	MWCNT	OH	1	1.06	30	50	0.5–2	>95	<1.5
6	MWCNT	COOH	0	0.73	30	50	0.5–2	>95	<1.5
7	MWCNT	COOH	3	4	10	20	0.5–2	>95	<1.5
8	SWCNT	COOH	0	2.73	1	4	0.5–2	>95	<1.5
9	MWCNT	COOH	0	3.86	1	8	0.5–2	>95	<1.5

^a *W_i* (%) = functional groups (OH, COOH)/carbon atoms ratio (%). The properties of the *i*th CNT are *D_i* = outer diameter, *L_i* = length, *P_i* = purity, *C_i* = electric conductivity.

Animal Welfare

Male *Wistar* rats (4 months old; approximately 150 g) received food and water ad libitum. They were kept in plastic cages with wire tops in a light-controlled room (12:12 h light–dark cycle) at 22 ± 3 °C before starting the study in accordance with the animal care and experimental procedures based on the Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes; these procedures were also approved by the Institutional Animal Care and Use Committee of the School of Pharmaceutical Sciences of Ribeirão Preto (CEUA-FCFRP) (license and registration number: 01.0263.2014).

Isolation of Rat Liver Mitochondria (RLM)

Mitochondria were isolated by standard differential centrifugation.(51) Male *Wistar* rats weighing approximately 200 g were euthanized by decapitation; livers (10–15 g) were immediately removed, sliced in a medium (50 mL) consisting of 250 mM sucrose, 1 mM ethylene glycol-bis(β -aminoethyl)-*N,N,N',N'*-tetraacetic acid (EGTA), and 10 mM HEPES-KOH, pH 7.2, and homogenized three times for 15 s at 1 min intervals using a Potter–Elvehjem homogenizer. Homogenates were centrifuged (2500 rpm, 5 min), and the resulting supernatant was further centrifuged (10 500 rpm, 10 min). Pellets were then suspended in a medium (10 mL) consisting of

250 mM sucrose, 0.3 mM EGTA, and 10 mM HEPES-KOH, pH 7.2, and centrifuged (6000 rpm, 15 min). The final mitochondrial pellet was suspended in a medium (1 mL) consisting of 250 mM sucrose and 10 mM HEPES-KOH, pH 7.2, and used within 3 h. Mitochondrial protein contents were determined by the Biuret reaction.

Standard Incubation Procedure

The isolated mitochondria were energized with 5 mM potassium succinate (plus 2.5 μ M rotenone) in a standard incubation medium consisting of 125 mM sucrose, 65 mM KCl, 2 mM inorganic phosphate (K_2HPO_4), and 10 mM HEPES-KOH pH 7.4 at 30 °C.

Continuous-Monitoring Mitochondrial Respiration Assays

The oxygen consumption in mitochondrial suspensions was polarographically determined with a Clark-type electrode. Clark electrodes have platinum cathodes and silver chloride anodes, which are connected by a salt bridge and covered by an oxygen-permeable membrane. As oxygen diffuses across the membrane, it is reduced by a fixed voltage between the cathode and anode that generates current in proportion to the concentration of oxygen in solution. By calibrating the voltage with known oxygen concentrations, it is possible to measure the rate of oxygen consumption in a medium containing actively respiring mitochondria. Since reduction of oxygen is a critical step in the process of mitochondrial electron transport and ATP synthesis, the measurement of mitochondrial oxygen consumption provides a convenient way to assess mitochondrial function.

To this end, two different set of tools were used: (1) Oxygraph System (Hansatech Instruments Ltd., Norfolk, UK) and (2) Oroboros Instruments (Oxygraph-2k). Both methodologies were applied in a 2 mL glass chamber equipped with a magnetic stirrer. Rat liver isolated mitochondria (1 mg protein/mL) were energized with 5 mM potassium succinate (plus 2.5 μ M rotenone) in a standard incubation medium consisting of 125 mM sucrose, 65 mM KCl, 2 mM inorganic phosphate (K_2HPO_4), and 10 mM HEPES-KOH pH 7.4 at 20 °C in a standard respiration medium. The experimental approach was calibrated using the oxygen content of an air-saturated medium.⁵¹

Theoretical Section

Raman Spectra Transform with Markov–Shannon Entropy Invariants

The current work proposes a new type of Raman spectra transform, similar to the Fourier transformation. This transform converts the Raman spectra values to character sequences and the corresponding star graphs (SGs) are constructed using S2SNet tool.⁵² A star graph is a special type of tree with N vertices, where one has $N - 1$ degrees of freedom and the remaining $N - 1$ vertices have a single degree of freedom.⁵³ In the case of protein sequences, the graph is built by adding all the amino acids into 23 possible branches (“rays” corresponding to the types of amino acids). The star center is a dummy node.⁵⁴

Thus, the Raman spectra were divided into intervals of 100 units, from 0 to 1800. As a result, the maximum number of star graph branches is 18 and corresponds to characters “a” to “r”. The star graph connectivity of the transformed Raman spectra provides the information needed to calculate the Shannon entropies (Sh_k , $k = 0-5$) for nonembedded and embedded SG (Sh and Sh_e). The transformation of the Raman spectra of CNT1 into Shannon entropies is shown in Figure 3.

The calculation of Sh invariants is based on matrices and vectors of the SG. To calculate Sh, the starting point is the SG connectivity matrix \mathbf{M} (dimension of $n = \text{number of nodes}$; $M_{ij} = 1$ if the nodes $i-j$ are connected; $0 = \text{if nodes } i-j \text{ are not connected}$). In the case of nonembedded SG, the connectivity is generated by the SG rule (each node is placed into a specific branch). In the case of embedded SG, in addition to the nonembedded connectivity, the sequence connectivity is added. S2SNet has been used to calculate nonembedded and embedded Sh for each CNT Raman spectrum with the following parameters: no weights for the nodes, Markov normalization, $k = 0-5$. The formula of Sh invariants is described in eq 1, and the details about the SG Shannon entropy formulation are presented in ref 52.

$$\text{Sh}_k = - \sum_{i=0}^n {}^k p_i \log({}^k p_i) \quad (1)$$

${}^k p_i$ = elements of the p vector obtained by multiplying the k powered Markov normalized matrix ($n \times n$) and a vector ($n \times 1$) with elements of $1/n$. The calculated values for all the CNTs are in the Figshare file “ds3.info.xlsx”, sheet “RAMAN entropies”.⁵⁵

Theoretical Details of the Nano-PT-QSPR Models

The general-purpose PT model for multiple boundaries has been proposed for chemoinformatics problems.²⁹ This study extends this theory to PT-QSPR models that will be able to study the effect of different CNTs on the mitochondrial respiration (oxygen consumption) under different experimental conditions. The general equation of the nano-PT-QSPR model is presented in eq 2.

$$E3_{pred} = e_0 + a_0 E3_{exp} + g_0 t + \sum_{k=0}^5 a_k d\text{Sh}_k + \sum_{k=0}^5 b_k d\text{She}_k \quad (2)$$

$E3_{pred}$ represents the predicted mitochondrial oxygen consumption. $E3_{exp}$ is the expected value of E3 in the set of three simultaneous experimental conditions (SEC). $d\text{Sh}_k$ and $d\text{She}_k$ are the moving averages of nonembedded and embedded Shannon entropies of the CNT Raman SG transform (differences between Sh_k or She_k and their averages obtained under different experimental conditions). The coefficients e_0 , g_0 , a_0 , a_k , and b_k are the equation optimal coefficients.

Model Data Set

The experimental data for the mitochondrial oxygen consumption (E3) in the presence of CNTs are available as a Figshare repository⁵⁵ and contain 16 335 cases (E3 measurements) with the following variables (data columns): E3, CNT label (CNT1-9), CNT type (CNT_type), type of CNT chemical modification (Function_type), type of the solvent (Solvent), and time (t). Thus, E3 was measured under three types of experimental conditions (c): CNT_type, Function_type (chemical modification of CNT), and Solvent. The codes used for CNTtypes are MWCNT, SW+DWCNT, and SWCNT. CNT type was set as 0 when the assay is a control assay with a blank solution with CNT concentration equal to 0. The values of Solvent condition are 0 and DMSO. The CNT Function types are 0 (none), COOH and OH. All the average values of the Shannon entropies under these experimental conditions are presented step by step in Figshare repository.⁵⁵ The final data set used to find the best prediction model is made up of 16 335 cases and 35 input features. The number of cases (N) is the result of $\text{NoCNTs} \times \text{TimePeriods} + \text{Replicates} + \text{CART.Blanks}$, where $\text{NoCNTs} = 9$ (different nanotubes) and $\text{TimePeriods} = 1485$ (min). Thus, $\text{NoCNTs} \times \text{TimePeriods} = 9 \times 1485 = 13\,365$. $\text{CART.Blanks} = 1485$ (experiments with

carboxyatractyloside (or CART as classical inhibitor of state V3 of respiration-ADP dependent at 1485 periods) were added. The remaining number of cases are replicates or blanks with H2O as solvent (see ds3.info.xlsx, sheet “Experimental Data”⁵⁵).

The following steps were used to generate the final data set (see Figure 4 and Figshare files⁵⁵):

- (1) Calculation of 11 nonembedded and embedded SG Shannon entropies for each CNT using S2SNet (Sh_k and She_k): $Sh_0, Sh_1, Sh_2, Sh_3, Sh_4, Sh_5, She_1, She_2, She_3, She_4, She_5$ (the zero-value descriptors have been excluded, CNT = 9 different CNTs)
- (2) Calculation of mean values for each Sh under an experimental conditions such as CNT_type = type of CNT, Solvent = type of solvent, Function_type = type of CNT chemical modification 11 CNTtype Sh_k /CNTtype She_k , 11 FuncType Sh_k /FuncType She_k , 11 Solvent Sh_k /Solvent She_k ; see “Experimental condition Means” in ds3.info.xlsx⁵⁵)
- (3) Calculation of 33 moving averages (MAs) between the original Raman spectra SG entropies (from step 1) and their averages obtained under the experimental conditions (from step 2) and time (for each type of CNT); the resulting values have the following notation in the Figshare files: 11 MA.FuncType. Sh_k /MA.CNTtype. She_k , 11 MA.CN.FuncType. Sh_k /MA.FuncType. She_k , 11 MA.Solvent. Sh_k /MA.Solvent. She_k
- (4) Calculation of the expected values of E3 ($^{SEC}E3_{exp}$) in a set of three experimental conditions (CNT type, solvent type, and CNT chemical function type; see Experimental condition Means in ds3.info.xlsx⁵⁵)
- (5) The final data set of 35 features was only made up of the experimental values (E3), the expected value of E3 in a set of experimental conditions from step 4 ($E3_{exp}$), time (t) and the 33 moving averages of Sh/She under all the experimental conditions from step 3 (MA.CNTtype. Sh_k /MA.CNTtype. She_k , MA.CN.FuncType. Sh_k /MA.FuncType. She_k , MA.Solvent. Sh_k /MA.Solvent. She_k)

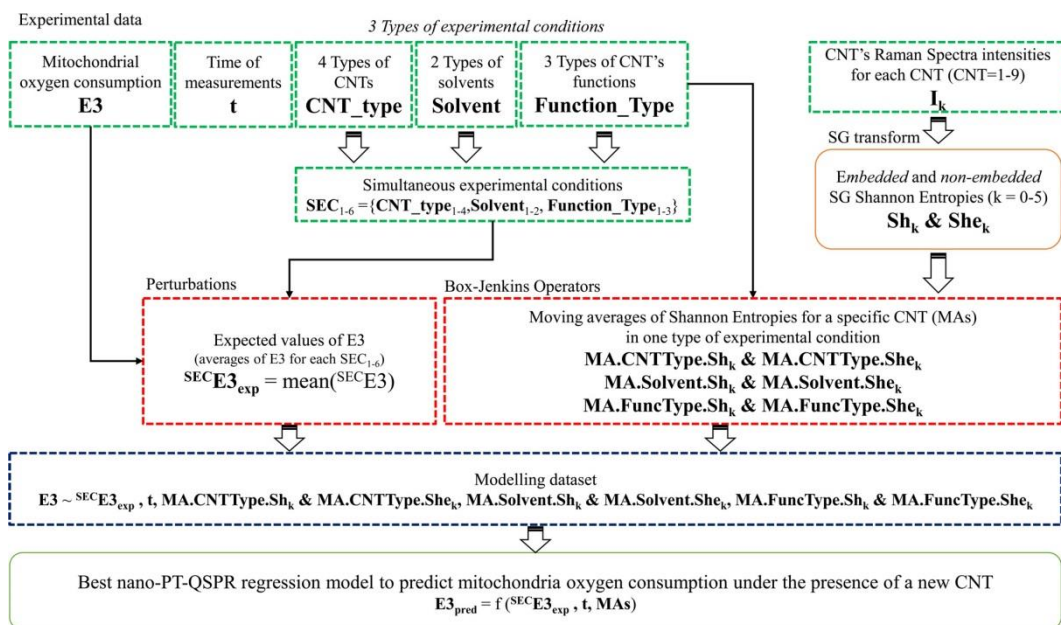


Figure 4. Methodology flowchart for nano-PT-QSPR models for mitochondrial oxygen consumption in the presence of CNTs.

Therefore, the entire flow could be summarize in few ideas:

- The study is searching for the best regression model that can predict values mitochondrial oxygen consumption (E3) measured in specific experimental conditions when different CNTs are present.
- The initial data to generate this model is composed on experimental E3 values (output/predicted variable), time of the E3 measurement, three types of experimental conditions (CNT type, solvent type, and chemical function type of the CNTs), and the Raman spectra of the CNTs.
- The current model is based on the perturbation theory, and it considers that E3 measured in three simultaneous experimental conditions is the sum of the expected value of E3 ($E3_{exp}$) in these conditions and some perturbations around this value ($E3 = E3_{exp} + \text{perturbations}$).
- Therefore, $E3_{exp}$ are obtained as mean values of E3 for three simultaneous experimental conditions (SEC). There are possible only six combinations of the three types of experimental conditions (CNT type and solvent type and chemical function type of the CNTs). Thus, $E3_{exp}$ values are constant features for the model.
- The perturbations of E3 around $E3_{exp}$ consists in a series of moving averages (MAs) of the SG Shannon entropies for the transformed Raman spectra for each CNT type and for one type of experimental condition (CNT type/solvent type/chemical function type of the CNTs). Therefore, the perturbations of E3 are MAs (Box–Jenkins operators) of SG Shannon entropies for specific experimental conditions (MA = difference of the entropy with the mean entropies in specific experimental condition).
- Because the values of E3 are measured at different times (time series), the time variable was added to the model (E3 is not constant in time).
- In conclusion, the prediction of E3 is made using SG Shannon entropy MAs as perturbations around the $E3_{exp}$ (expected values of E3) and time as a time series of data.

The advantages of the PT-QSAR technique over the conventional QSAR are the following:

- The PT method uses the perturbations of the classic QSAR features around the average values of the same features in different experimental conditions (moving averages). Therefore, the PT methodology is able to compare relative values of the model features, which are much smaller compared to the absolute values of these features. PT-QSAR is based on the perturbations of the features and not on the feature values, offering information more detailed than the traditional QSAR. In the latter, if the feature values are high, the small differences between them may be undetectable by the statistical or machine learning techniques.
- The PT method averages feature values using experimental conditions (CNT type, solvent type, and CNT chemical function type). Thus, not only is molecular information included in the QSAR model but also the information about specific experimental conditions for the observed model output. The experimental conditions affect the observed variables, and therefore, including this dependent information in the future QSAR model is a big advantage.
- The introduction of the expected value of the output variable ($E3_{exp}$) is additional information added to the QSAR model. $E3_{exp}$ is the average value of the output E3 measured in a combination of experimental conditions (not only one experimental condition). This average value adds information about the measurement experiments. Thus, the predicted output $E3_{pred}$ is calculated similarly to a perturbation around the $E3_{exp}$ values for specific experimental conditions (see eq 2). The perturbations are represented by the MA of the CNT descriptors in different experimental conditions. MAs are calculated using an experimental condition, and $E3_{exp}$ was obtained using the combination of three experimental conditions (CNT type, solvent type, and CNT chemical function type).

The use of the best regression model for the prediction of E3 values (mitochondrial oxygen consumption) in the presence of a new CNT (with a new Raman spectra) consists in providing the input features for the final model such as $E3_{\text{exp}}$, time, and MAs of CNT's Shannon entropies for transformed Raman spectra: $E3_{\text{pred}} = f(\text{SEC}E3_{\text{exp}}, t, \text{MAs})$, where SEC = simultaneous experimental conditions and MAs are calculated for all types of CNTs. Thus, it is very important to specify specific experimental condition and time for the prediction of E3 values.

Let us consider the prediction of the mitochondrial oxygen consumption ($E3_{\text{pred}}$) for specific values of the experimental conditions (CNT type and solvent type and chemical function type of the CNTs) and at a specific time moment after a new CNT_x was added. Thus, the prediction is made for specific experimental conditions and, therefore, a specific value of $\text{SEC}E3_{\text{exp}}$ will be used (from the six values calculated for building the regression model). It is not necessary to calculate any other value of $\text{SEC}E3_{\text{exp}}$ because the model will evaluate the $E3_{\text{pred}}$ value based on the $E3_{\text{exp}}$ values obtained with the model data set. The second input feature is the time of the CNT presence in mitochondria.

After picking the time and $E3_{\text{exp}}$ is known for specific experimental conditions, only the MAs should be calculated for the new prediction. First of all, the Raman spectra of the new CNT will be transformed into SG Shannon entropies (Sh). From the model building, the averages values of the CNTs in specific experimental conditions are known, and, therefore, the MAs of the new CNT_x will be calculated as a difference between the new Sh values and the model's averages of Sh of all the data set CNTs. Thus, the model is providing the $E3_{\text{exp}}$ values and Sh means for specific experimental conditions, but the user will choose what values to use depending on experimental conditions for the new E3 prediction.

Regression Predictors

The raw data set was normalized and the training and test subsets were obtained using 10 splits: 75% training set (train) and 25% test set (test). The raw and normalized data sets are available online with the R script for normalization and data splitting.⁵⁵ The batchRRegrs tool was used to find the best regression nano-PT-QSPR model. The models were selected using the R_{ts}^2 values (*R*-squared) and the RMSE_{ts} (root-mean-square error) corresponding values for test subset.

RRegrs is an R integrated framework that provides ten linear and nonlinear regression models.^{45,46} Due to the computational limitations, only four RRegrs methods were used: multiple linear (LM), partial least squares (PLS)⁵⁶ and neural networks (NN) regressions⁵⁷ and random forest (RF).⁵⁸ Generally, default values of parameters were used. In the case of NN and RF, the variation of the method parameters was studied. The RRegrs call is not prepared for big data sets and for some parameter variations. Thus, a modified version of RRegrs (batchRRegrs: <https://github.com/cafernandezlo/batchRRegrs>) was used on the BioCAI HPC platform from University of A Coruna (Spain). This version of RRegrs saves the R model objects and it leaves the door open for any type of extra calculations or graphical plots for the regression model. As an adaptation for large data sets, several features are missing in batchRRegrs: there is only one split (the user runs create each split), there is no Y-randomization, there are no output figures as PDFs and there is no automatic selection of the best model. Therefore, the current methodology is an incomplete RRegrs flow, due to the mission of the Y-randomization. The criteria to find the best model are the same as for RRegrs: maximum R_{ts}^2 and minimum RMSE_{ts}. The plots for the current work were obtained with R scripts. The best regression model which predicts mitochondrial oxygen consumption in the presence of CNTs is available online⁵⁵ and it can be used for future predictions.

RESULTS AND DISCUSSION

In principle, several biochemical *in vitro* assays may be used with different respiratory substrates, cofactors, and inhibitors in the polarographic evaluation of the effect of CNT family on the mitochondrial oxygen consumption. The rate of oxygen consumption can be measured directly using a Clark-type electrode, which consists of a probe with an exposed platinum cathode and a silver anode. When the anode and cathode are polarized, the current produced is directly proportional to the partial pressure of oxygen. The respiratory biochemical reactions are connected via an electrolyte solution, such as KCl. The cathode is typically covered by an oxygen-permeable membrane, such as a polypropylene membrane, to exclude contaminating species, ions, or samples that might interfere with the reaction.⁵⁹

In this work, two different polarographic experiments were performed to evaluate the effects of the CNT family on the mitochondrial respiration by exposure conditions of 5 µg/mL.

In general terms, the results showed that the entire CNT family tested does not inhibit (or affect) the profiles of oxygen consumption increment after ADP addition (state V3 of ADP-dependent respiration) compared to the untreated mitochondrial control (black line), no significant differences being detected ($p > 0.05$) in this case. However, for CNT-1 (pristine MWCNT) a moderate ability to inhibit the state V3 of mitochondrial ADP-dependent respiration (orange line), compared to their similar oxidized CNT (hydroxylated-CNTs CNT2–CNT5 and carboxylated-CNTs CNT6–CNT9) was observed. The mitochondrial V3 respiratory inhibition by CNT-1 is not significant (ns: $p > 0.05$) when compared to the untreated mitochondrial control (black line) under the evaluated experimental conditions. It does not affect the oxidative phosphorylation by ATP synthase, which depends on the ADP transport by the ADP/ATP mitochondrial carrier between cytosol and the mitochondrial matrix under physiological normoxic conditions, the *in vitro* results suggesting lower inhibition response for isolated rat liver mitochondria based on mitochondrial V3 respiratory inhibition. Generally, a low inhibition for each member of the CNT family was observed, represented by colored lines, related to untreated mitochondrial control (black line) after the addition of CCCP 1 µM (uncoupling agent of the mitochondrial oxidative phosphorylation) in the state V4 of respiration and V_{cccp} (both ADP independent) shown in Figure 5. Intriguingly, it was observed that the mitochondrial respiration rate versus time of V4 (or V_{cccp}) stages (ADP independent) for CNT-1 (pristine MWCNT) (orange line) had the smallest decline, with significant statistical differences (** $p < 0.05$) compared to their similarly oxidized CNTs (CNT2–CNT9) (the remaining colored lines) and also to the untreated mitochondrial control (black line), despite the presence of protonophore CCCP, a classic mitochondrial uncoupling agent, which represents the maximum rate of mitochondrial oxygen consumption used as control at the end of all the measurements ($t = 500$ s).

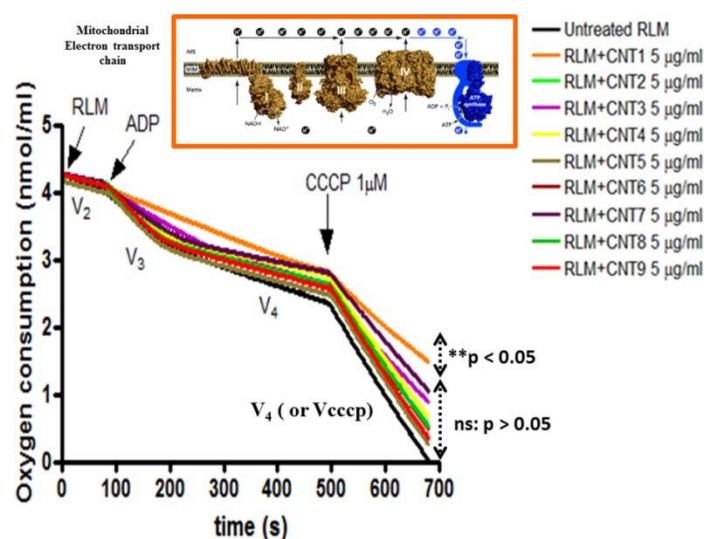


Figure 5. Profiles of mitochondrial oxygen consumption of isolated rat-liver mitochondria show the different states of mitochondrial respiration on the mitochondrial respiratory chain: V_2 state (basal respiration), V_3 state (ADP-dependent respiration), V_4 and V_{cccp} states (both ADP-independent respiration) in untreated-rat liver mitochondria or untreated-RLM (black line) and after the CNT exposure at 5 $\mu\text{g}/\text{mL}$ (remaining colors). These results are representative of three experiments by using the two polarographic assessments of mitochondrial respiration performed with the Oxygraph System Hansatech Instruments and Oroboros Instruments (Oxygraph-2k). $**p < 0.05$ is used to represent the statistical differences between CNT-1 (pristine MWCNT) (orange line) and untreated RLM (black line).

A physiological explanation is feasible due to the fact that carbon nanotubes can pass through phospholipid bilayers that make up the mitochondrial inner membrane. More specifically, they are accumulated similarly too many cationic amphiphilic drugs or in some cases when degraded by the oxido-reduction mechanisms that occur in the mitochondrial matrix. The effects can be observed in the interruption of the flow of electrons, decreasing the proton gradient associated with low oxygen consumption rate, uncoupling of oxidative phosphorylation and reduction of ATP synthesis, depending on the physical–chemical nature of the mitotoxic agent involved. According to this idea, it is well-known that the presence of defects such as pentagons, heptagons, vacancies, or metallic dopant are found to modify drastically the electronic properties of these nanosystems (carbon nanotubes) and, in the same vein, their properties of interaction with the biological systems, which can be expressed through biocompatibility and/or toxicity responses. In this sense, the intensity of the D band of CNT tested regarding the Raman spectra with peak at approximately 1350 cm^{-1} is commonly associated with the presence of the aforementioned topological modifications in the carbon lattice of CNTs.

Thus, the presence of the CNT's oxidized groups ($-\text{OH}$, $-\text{COOH}$) with a peak at 1350 cm^{-1} can prevent inhibition of the V_3 state (ADP dependent) or uncoupling effects on mitochondrial respiration state (V_4), maybe associated with the oxidized CNT ability as free oxygen-radical scavenging. Since these oxygen free radicals were spontaneously formed in the mitochondrial complex I and III, perturbation can be induced in the oxidative phosphorylation of the ADP and Pi by the mitochondrial complex V wrapping. In most cases inhibition and/or uncoupling effects on mitochondrial respiration, similar to the effects observed due to the addition of CCCP or V_4 (V_{cccp}) in all previous states of mitochondrial respiration (V_2 , V_3) were not observed in the present study. Raman spectroscopy D band characterized the level of the CNT oxidation, and it has been recognized that CNT-covalent functionalization can generate different Stones–Wales defects formed by rotating a C–C bond by 90° in the sidewalls of the CNT structure with loss of sp^2 -carbon hybridization or rehybridization (ability of the carbon atom to hybridize between sp^2 and sp^3). According to the theoretical results obtained by Galano et al. using semiempirical methods

such as density functional theory (DFT), moderate, or high oxidation on the carbon lattice of CNT increases their oxygen free-radical scavenging activity and has no impact on mitochondrial respiration states by capturing the oxygen radical species formed in the mitochondrial complex I and III. These biochemical effects, which may involve C atoms with dangling bonds, cause a larger increase of CNT reactivity toward free radicals than the Stones–Wales and vacancy defects without C atoms with dangling bonds.⁶⁰⁻⁶⁴ Interestingly, it was demonstrated that pristine CNT (MWCNT-1) in aqueous suspension does not generate oxygen-free radicals.

On the contrary, it was observed that MWCNTs exhibit a remarkable free radical scavenging ability, when in contact with an external source of oxygen-free radicals (hydroxyl or superoxide radicals), suggesting that the multiwalls of MWCNTs can act as oxygen-free radical sponges, and therefore, they do not affect the mitochondrial respiration by removing potential oxygen radicals from the carbon lattice. Recent studies have shown that pristine SWCNT are more reactive to induce mitochondrial damage than their oxidized forms SWCNT-COOH. Our hypothesis is that OH and COOH groups from CNT family could interact with the metallic centers of Fe³⁺ and Cu²⁺ present in the respiratory complexes preventing redox change to form Fe²⁺ and Cu¹⁺, respectively. These compounds present less ability to reduce oxygen to water and promote abnormalities in mitochondrial bioenergetic processes. However, it should be emphasized the good relationship between the biocompatibility and toxicity for the oxidized-CNTs, compared to their similar pristine-CNTs (CNT-1), whose mitotoxicity potential is greater.

In this regard, recent experiments using pristine and oxidized-CNT porin have shown the high potential as synthetic analogues of biological membrane channels with high efficiency and selectivity for transporting ions and molecules^{53,54} like natural substrates (ADP, ATP). According to this idea, the tested CNT family can spontaneously insert lipid bilayers into the cellular and mitochondrial membrane to form CNT channels that exhibit a unitary conductance of 70–100 pS under physiological conditions.⁶⁵ At the same time, the attenuation of the negative charge of (COO⁻) moieties of carboxylated CNT to form COOH moieties by the H⁺ protons present in the intermembrane mitochondrial space, may allow the electrostatic ADP³⁻ anions passage without inducing uncoupling effects on the mitochondrial respiratory function. In addition, Tunuguntla⁶⁶ showed through molecular dynamics simulations that carbon nanotube-porins can favor the ultrafast proton transport in sub-1 nm diameter and that the transport rates in these narrow nanotube pores can also exceed those of biological channels (as respiratory complexes that form the mitochondrial electron transport chain) due to confinement in hydrophobic nanochannels. In this sense, it could be hypothesized that CNTs with diameter >1 nm, such as those used in the present study, could have lower transport rates for the H⁺ influx, preventing possible mitochondrial uncoupling effects.

The structural determinants of mitochondrial mechanisms of pristine and functionalized (oxidized) CNTs remain poorly understood at present. Our suggestion is that CNT-COOH and CNT-OH should be more biocompatible compared to the pristine MWCNTs based on the influence on the respiratory mechanism under the experimental conditions tested and according to the time of exposure to the CNT assayed.

In addition, the variation in time (*t*) of the mitochondrial oxygen consumption in the presence of specific CNTs was modeled using nano-PT-QSPR methodology. For each type of nanotube, the corresponding star graph Shannon entropies of the Raman spectra (Sh₀₋₅/She₀₋₅; e = embedded SG) were calculated. For each set of experimental conditions, the corresponding expected values for the E (E_{exp}) and the moving averages of the SG descriptors (MA.[experimental conditions].Sh₀₋₅, MA.[experimental conditions].She₀₋₅) were calculated. The final data set has 35 features: E_{3 exp}, *t*, MA.CNTType.Sh₀₋₅, MA.FuncType.Sh₀₋₅, MA.Solvent.Sh₀₋₅, MA.CNTType.She₀₋₅, MA.FuncType.She₀₋₅, and MA.Solvent.She₀₋₅. The output variable is E₃, and there are 16 335 cases. A modified version of the RRegrs tool was used on an HPC cluster in order to test four types of regression methods. The purpose was to find the best prediction model for mitochondrial oxygen consumption in the presence of specific CNTs. Thus, linear and nonlinear methods were

used for the feature pool: LM, PLS, NN, and RF (10 different splits and repeated cross-validation parameter). All the results are based on 35 features (pool). Table 2 shows the minimum, maximum, and mean values for the $R_{tr}^2/RMSE_{tr}$ and $R_{ts}^2/RMSE_{ts}$. The small differences between the training and test statistics demonstrated that the models are not overfitted. The statistics for 10 splits for each method could be found in the Figshare project (ds3.models.xlsx).⁵⁵ The distribution of R^2 and RMSE (tr and ts) are presented in the same file as graphics.

Table 2. Training and Test R^2 and RMSE (Mean, Minimum, and Maximum Values for 10 Splits) Using RRegrs Regression Methods to Predict Mitochondrial Respiration Modifications Due to CNTs

method	R^2	mean	min	max	RMSE	mean	min	max
LM	R_{tr}^2	0.865	0.864	0.866	RMSE _{tr}	0.0915	0.0910	0.0922
	R_{ts}^2	0.864	0.861	0.868	RMSE _{ts}	0.0913	0.0891	0.0929
PLS	R_{tr}^2	0.864	0.863	0.865	RMSE _{tr}	0.0917	0.0912	0.0925
	R_{ts}^2	0.864	0.860	0.867	RMSE _{ts}	0.0916	0.0893	0.0931
NN	R_{tr}^2	0.987	0.986	0.987	RMSE _{tr}	0.0286	0.0283	0.0292
	R_{ts}^2	0.988	0.987	0.990	RMSE _{ts}	0.0271	0.0249	0.0284
RF	R_{tr}^2	0.998	0.998	0.998	RMSE _{tr}	0.0109	0.0102	0.0118
	R_{ts}^2	0.998	0.997	0.999	RMSE _{ts}	0.0108	0.0068	0.0133

The results show the possibility to predict the effect of nanotubes on the mitochondrial respiration using a 35-feature linear model (LM) with a mean R_{ts}^2 value of 0.864 and a mean $RMSE_{ts}$ value of 0.0913. The MAs of the star graph Shannon entropies under different experimental conditions provide enough information for this linear regression predictor. PLS obtain results similar to those of LN, without removing any feature from the final model. Two attempts of feature selection were performed using Lasso (14 features, $R_{ts}^2 = 0.865$, $RMSE_{ts} = 0.0921$) and elastic net (23 features, $R_{ts}^2 = 0.864$, $RMSE_{ts} = 0.0926$) regression methods from RRegrs, but without improvements when compared to LM/PLS. The selected features included E_{exp} , t , and feature moving averages. The significant increase in the regression performance is observed when nonlinear regression methods such as NN and RF have been used. Thus, NN provides a mean R_{ts}^2 of 0.988 and a mean $RMSE_{ts}$ value of 0.0271. Thus, NN has an $RMSE_{ts}$ value over three times smaller than LM. Figure 6 shows the study of the optimal NN parameters for the final model: number of hidden neurons = [1, 5, 10, 15, 20, 50] and weight decay = [0, 0.00001, 0.005, 0.001, 0.1]. The best NN model (split 5)⁵⁵ is characterized by a single hidden layer with 10 neurons, a structure of 35–10–1 (35 inputs, 10 neurons in one hidden layer, 1 output), a weight decay of 0.001, and R_{ts}^2 of 0.990 and $RMSE_{ts}$ of 0.0249.

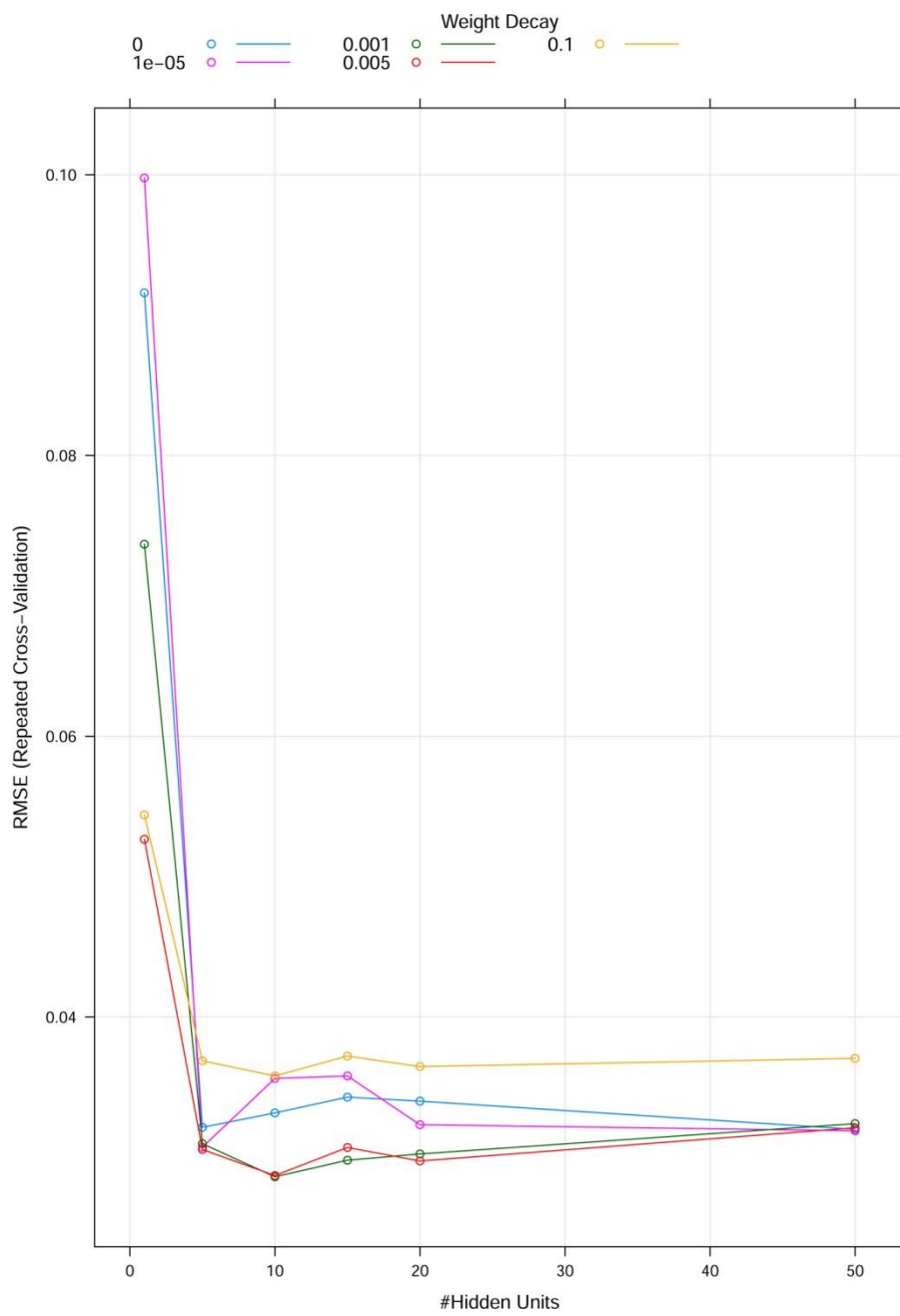


Figure 7. Parameter study for RF regression models.

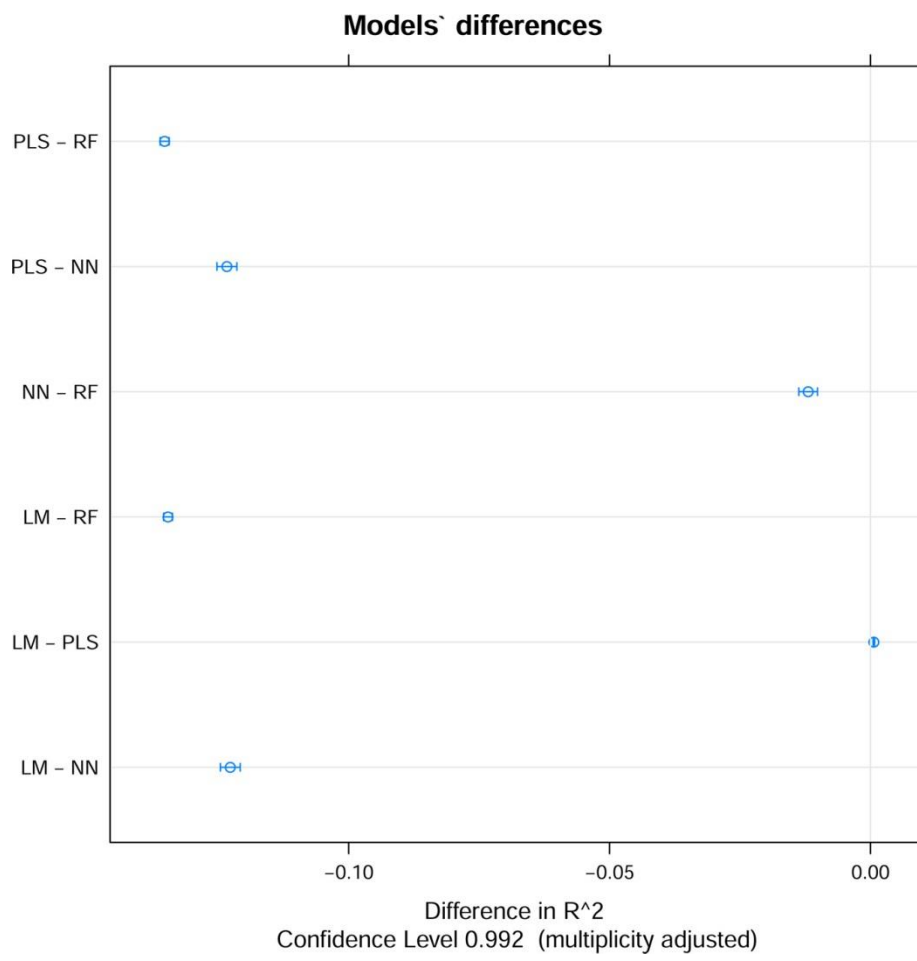


Figure 8. RRegrs pairwise model comparisons of R_{is}^2 .

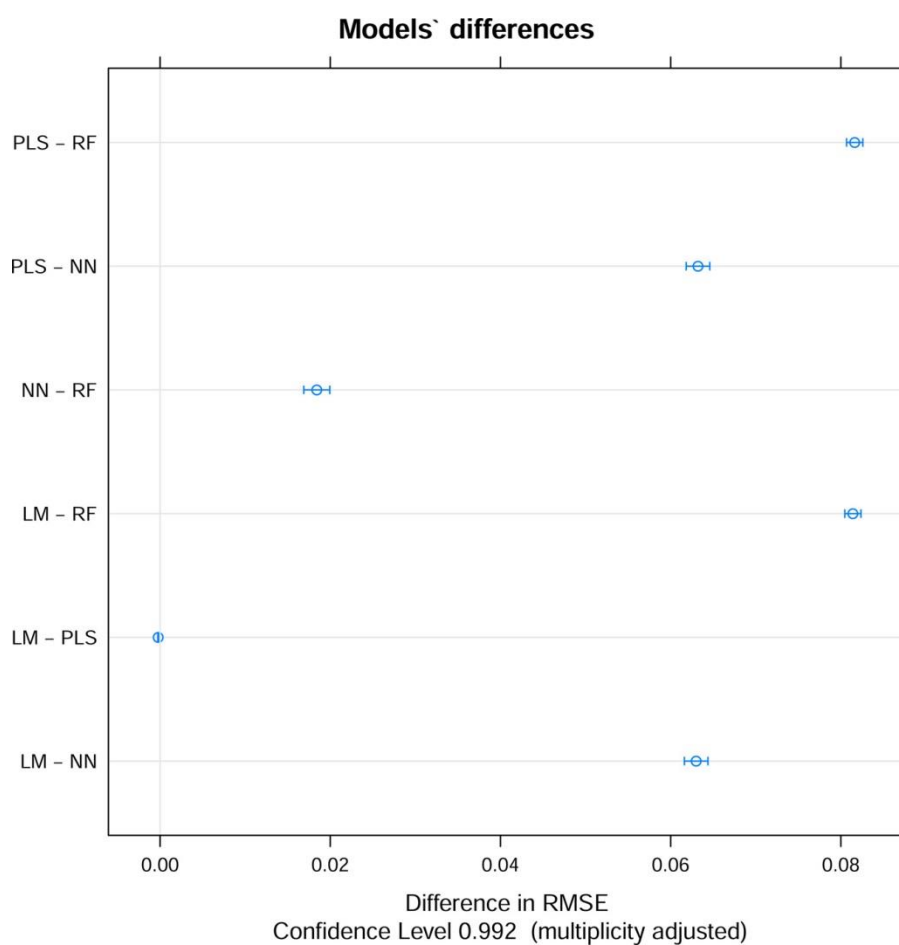


Figure 9. RRegrs pairwise model comparisons of $RMSE_{is}$.

Any of these models could be used for prediction due to the similar statistics. In order to help the users, one model has been exported as an R object (“RF.details.5.csv.split.1.rf.model.RData” from the Figshare repository⁵⁵) in order to be directly used for predictions of E3 values for a new CNT (without the necessity of recalculate the RF model). In addition, this model can be implemented in other computational tools. All the models can be obtained with the Figshare data set, batchRRegrs and the seed as the split number. The prediction of a new E3 values consists in the following steps: (1) measurement or achievement of the Raman spectra of the new CNT; (2) calculation of the SG Shannon entropies for the Raman spectra (Sh_k/She_k); (3) calculation of the new moving averages of Sh for each type of experimental condition, such as the differences between the new descriptors and the mean of Sh under different conditions from the model; (4) use of expected values from the model for a specific combination of experimental conditions (limited to the combinations in the model); (5) setting time (t) values; (6) use of the model to predict the mitochondrial oxygen consumption ($E3_{pred}$) in the presence of the new CNT, for a specific experimental condition and in a specific time scale.

CONCLUSIONS

The respiratory functional assays showed that CNT Raman spectroscopy structural nanodescriptors could be useful to evaluate toxicity criteria on mitochondrial respiration for state V3 of respiration and oxidative phosphorylation. Experimental biochemical results pointed out that encrypted information in the Raman spectra of CNT structure could be associated with canonical cellular bioenergetic mechanism to be employed for making regulatory decisions in nanotoxicology and increasing the potential biomedical application of new carbon nanomaterials. It is the first time when a star graph transform of the CNT Raman spectra is proposed in order to obtain the raw information for a nano-PT-QSPR model. Thus, Box–Jenkins and PT operators of star graph Shannon entropies under different experimental conditions and time scales were used to find the best regression predictor. A modified version of the RRegrs tool was used to test four linear and nonlinear methods such as LN, PLS, NN, and RF. The RF method provided the best models to predict the toxicity of CNTs in mitochondrial respiration with R^2 of 0.998–0.999 and RMSE of 0.0068–0.0133 for the training and test subset. As a result, this work demonstrated the SG transform power when encoding Raman spectra information, similar to the SG transform of the blood proteome spectra in cancer or electroencephalogram in epilepsy and their applicability as prospective chemoinformatics tool in nanorisk assessment.

AUTHOR CONTRIBUTIONS

M.G. was involved in the Raman characterization and testing in functional biochemical respiration on isolated mitochondria. L.C.A. contributed to the polarography biochemical assays, the use of mitochondrial inhibitors, the interpretation of results, and the management of laboratory animals. Z.N. helped with the characterization of the CNTs. D.T.A.-S. was involved in the characterization of CNTs and helped with the CNT samples in the isolated mitochondria model. C.C. assisted in the interpretation of experimental results, taking into account the ethical principles regarding laboratory animals, infrastructure, and isolation techniques implemented to study mitochondria and bioenergetics. H.G.-D. applied the star graph transform of Raman spectra, and the perturbation theory and moving averages methodologies. J.M.V.-N. and C.R.M. used the machine learning techniques to find the best regression models and characterized the models employing R scripting. The manuscript was written through contributions of all authors.

The authors declare no competing financial interest.

ACKNOWLEDGMENT

M.G.-D. acknowledges a doctoral fellowship (Postgraduate Students Program PEC-PG No.062/2013) from Brazilian Agencies CAPES-CNPq. Furthermore, M.G.-D. would like to thank to Prof. J. M. Monserrat for his valuable supervision. This work was also supported by the General Directorate of Culture, Education and University Management of Xunta de Galicia (ref GRC2014/049) and the European Regional Development Fund (FEDER). In addition, the work was supported by the “Galician Network for Colorectal Cancer Research (REGICC)” (ref R2014/039), funded by Xunta de Galicia.

ABBREVIATIONS

CNT, carbon nanotube; SWCNT, single-walled carbon nanotube; MWCNT, multiwalled carbon nanotube; SG, star graph; LM, multiple linear regression; PLS, partial least squares regression; NN, neural networks regression; RF; random forest regression; RMSE, root-mean-square error; ATP, adenosine triphosphate; ADP, adenosine diphosphate; CCCP, carbonyl cyanide *m*-chlorophenyl hydrazine; CART, carboxyatractyloside; QSPR/QSAR, quantitative structure–property/activity relationships; PT, perturbation theory; DFT, density functional theory; EEG, electroencephalogram; RLM, rat liver mitochondria; S2SNet, sequence to star network (software); SEC, simultaneous experimental conditions; Mas, moving averages; Sh, Shannon entropy

REFERENCES

This article references 66 other publications.

- (1) Segawa, Y.; Yagi, A.; Matsui, K.; Itami, K. Design and Synthesis of Carbon Nanotube Segments *Angew. Chem., Int. Ed.* 2016, 55, 5136– 58 DOI: 10.1002/anie.201508384
- (2) Kroemer, G.; Reed, J. C. Mitochondrial control of cell death *Nat. Med.* 2000, 6, 513– 9 DOI: 10.1038/74994
- (3) Kroemer, G. The Mitochondrial Permeability Transition Pore Complex as a pharmacological target. An introduction *Curr. Med. Chem.* 2003, 10, 1469– 72 DOI: 10.2174/0929867033457232
- (4) Pacurari, M.; Lowe, K.; Tchounwou, P. B.; Kafoury, R. A Review on the Respiratory System Toxicity of Carbon Nanoparticles *Int. J. Environ. Res. Public Health* 2016, 13, 325 DOI: 10.3390/ijerph13030325
- (5) Liu, J. H.; Yang, S. T.; Wang, H.; Liu, Y. Advances in biodistribution study and tracing methodology of carbon nanotubes *J. Nanosci. Nanotechnol.* 2010, 10, 8469– 81 DOI: 10.1166/jnn.2010.2689
- (6) Li, Z.; Graham, B. H. Measurement of mitochondrial oxygen consumption using a Clark electrode *Methods Mol. Biol.* 2012, 837, 63– 72 DOI: 10.1007/978-1-61779-504-6_5
- (7) Wang, X.; Guo, J.; Chen, T.; Nie, H.; Wang, H.; Zang, J.; Cui, X.; Jia, G. Multi-walled carbon nanotubes induce apoptosis via mitochondrial pathway and scavenger receptor *Toxicol. In Vitro* 2012, 26, 799– 806 DOI: 10.1016/j.tiv.2012.05.010
- (8) Knobeloch, L. M.; Blondin, G. A.; Harkin, J. M. Use of submitochondrial particles for prediction of chemical toxicity in man *Bull. Environ. Contam. Toxicol.* 1990, 44, 661– 668 DOI: 10.1007/BF01701785
- (9) Hastings, J.; Jeliakova, N.; Owen, G.; Tsiliki, G.; Munteanu, C. R.; Steinbeck, C.; Willighagen, E. eNanoMapper: harnessing ontologies to enable data integration for nanomaterial risk assessment *J. Biomed Semantics* 2015, 6, 10 DOI: 10.1186/s13326-015-0005-5
- (10) Garcia-Fuentes, M.; Gonzalez-Diaz, H.; Csaba, N. Nanocarriers & drug delivery: rational design and applications *Curr. Top. Med. Chem.* 2014, 14, 1095– 6 DOI: 10.2174/1568026614666140329223843
- (11) Tantra, R.; Oksel, C.; Puzyn, T.; Wang, J.; Robinson, K. N.; Wang, X. Z.; Ma, C. Y.; Wilkins, T. Nano(Q)SAR: Challenges, pitfalls and perspectives *Nanotoxicology* 2015, 9, 636 DOI: 10.3109/17435390.2014.952698
- (12) Gajewicz, A.; Schaeublin, N.; Rasulev, B.; Hussain, S.; Leszczynska, D.; Puzyn, T.; Leszczynski, J. Towards understanding mechanisms governing cytotoxicity of metal oxides nanoparticles: Hints from nano-QSAR studies *Nanotoxicology* 2015, 9, 313 DOI: 10.3109/17435390.2014.930195
- (13) Kar, S.; Gajewicz, A.; Puzyn, T.; Roy, K. Nano-quantitative structure-activity relationship modeling using easily computable and interpretable descriptors for uptake of magnetofluorescent engineered nanoparticles in pancreatic cancer cells *Toxicol. In Vitro* 2014, 28, 600– 6 DOI: 10.1016/j.tiv.2013.12.018

- (14) Toropov, A. A.; Toropova, A. P.; Puzyn, T.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. QSAR as a random event: modeling of nanoparticles uptake in PaCa2 cancer cells *Chemosphere* 2013, 92, 31– 7 DOI: 10.1016/j.chemosphere.2013.03.012
- (15) Toropov, A. A.; Toropova, A. P.; Benfenati, E.; Gini, G.; Puzyn, T.; Leszczynska, D.; Leszczynski, J. Novel application of the CORAL software to model cytotoxicity of metal oxide nanoparticles to bacteria *Escherichia coli* *Chemosphere* 2012, 89, 1098– 102 DOI: 10.1016/j.chemosphere.2012.05.077
- (16) Puzyn, T.; Rasulev, B.; Gajewicz, A.; Hu, X.; Dasari, T. P.; Michalkova, A.; Hwang, H. M.; Toropov, A.; Leszczynska, D.; Leszczynski, J. Using nano-QSAR to predict the cytotoxicity of metal oxide nanoparticles *Nat. Nanotechnol.* 2011, 6, 175– 8 DOI: 10.1038/nnano.2011.10
- (17) Puzyn, T.; Leszczynska, D.; Leszczynski, J. Toward the development of "nano-QSARs": advances and challenges *Small* 2009, 5, 2494– 509 DOI: 10.1002/sml.200900179
- (18) Sizochenko, N.; Rasulev, B.; Gajewicz, A.; Kuz'min, V.; Puzyn, T.; Leszczynski, J. From basic physics to mechanisms of toxicity: the "liquid drop" approach applied to develop predictive classification models for toxicity of metal oxide nanoparticles *Nanoscale* 2014, 6, 13986– 93 DOI: 10.1039/C4NR03487B
- (19) Carbo-Dorca, R.; Besalu, E. Construction of coherent nano quantitative structure-properties relationships (nano-QSPR) models and catastrophe theory *SAR QSAR Environ. Res.* 2011, 22, 661– 5 DOI: 10.1080/1062936X.2011.623319
- (20) Shao, C. Y.; Chen, S. Z.; Su, B. H.; Tseng, Y. J.; Esposito, E. X.; Hopfinger, A. J. Dependence of QSAR models on the selection of trial descriptor sets: a demonstration using nanotoxicity endpoints of decorated nanotubes *J. Chem. Inf. Model.* 2013, 53, 142– 58 DOI: 10.1021/ci3005308
- (21) Toropova, A. P.; Toropov, A. A. Mutagenicity: QSAR - Quasi-QSAR - Nano-QSAR Mini-Rev. *Med. Chem.* 2015, 15, 608 DOI: 10.2174/1389557515666150219121652
- (22) Toropova, A. P.; Toropov, A. A.; Benfenati, E. A quasi-QSPR modelling for the photocatalytic decolourization rate constants and cellular viability (CV%) of nanoparticles by CORAL SAR *QSAR Environ. Res.* 2015, 26, 29– 40 DOI: 10.1080/1062936X.2014.984327
- (23) Oksel, C.; Ma, C. Y.; Wang, X. Z. Current situation on the availability of nanostructure-biological activity data SAR QSAR *Environ. Res.* 2015, 26, 79– 94 DOI: 10.1080/1062936X.2014.993702
- (24) Shahbazy, M.; Kompany-Zareh, M.; Najafpour, M. M. QSAR analysis for nano-sized layered manganese-calcium oxide in water oxidation: An application of chemometric methods in artificial photosynthesis *J. Photochem. Photobiol., B* 2015, 152, 146 DOI: 10.1016/j.jphotobiol.2014.12.020
- (25) Gajewicz, A.; Cronin, M. T.; Rasulev, B.; Leszczynski, J.; Puzyn, T. Novel approach for efficient predictions properties of large pool of nanomaterials based on limited set of species: nano-read-across *Nanotechnology* 2015, 26, 015701 DOI: 10.1088/0957-4484/26/1/015701
- (26) Toropova, A. P.; Toropov, A. A.; Rallo, R.; Leszczynska, D.; Leszczynski, J. Optimal descriptor as a translator of eclectic data into prediction of cytotoxicity for metal oxide nanoparticles under different conditions *Ecotoxicol. Environ. Saf.* 2015, 112, 39– 45 DOI: 10.1016/j.ecoenv.2014.10.003
- (27) Toropova, A. P.; Toropov, A. A.; Benfenati, E.; Korenstein, R.; Leszczynska, D.; Leszczynski, J. Optimal nano-descriptors as translators of eclectic data into prediction of the cell membrane damage by means of nano metal-oxides *Environ. Sci. Pollut. Res.* 2015, 22, 745– 57 DOI: 10.1007/s11356-014-3566-4
- (28) Burello, E.; Worth, A. P. QSAR modeling of nanomaterials *Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol.* 2011, 3, 298– 306 DOI: 10.1002/wnan.137
- (29) Gonzalez-Diaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M. General theory for multiple input-output perturbations in complex molecular systems. 1. Linear QSPR electronegativity models in physical, organic, and medicinal chemistry *Curr. Top. Med. Chem.* 2013, 13, 1713– 41 DOI: 10.2174/1568026611313140011
- (30) Gonzalez-Diaz, H.; Arrasate, S.; Juan, A. G.; Sotomayor, N.; Lete, E.; Speck-Planche, A.; Ruso, J. M.; Luan, F.; Cordeiro, M. N. Matrix trace operators: from spectral moments of molecular graphs and complex networks to perturbations in synthetic reactions, micelle nanoparticles, and drug ADME processes *Curr. Drug Metab.* 2014, 15, 470– 88 DOI: 10.2174/1389200215666140908101604
- (31) Vergara-Galicia, J.; Prado-Prado, F. J.; Gonzalez-Diaz, H. Galvez-Markov network transferability indices: review of classic theory and new model for perturbations in metabolic reactions *Curr. Drug Metab.* 2014, 15, 557– 64 DOI: 10.2174/1389200215666140605125827
- (32) Gonzalez-Diaz, H.; Perez-Montoto, L. G.; Ubeira, F. M. Model for vaccine design by prediction of B-epitopes of IEDB given perturbations in peptide sequence, in vivo process, experimental

- techniques, and source or host organisms *J. Immunol. Res.* 2014, 2014, 768515 DOI: 10.1155/2014/768515
- (33) Luan, F.; Kleandrova, V. V.; Gonzalez-Diaz, H.; Ruso, J. M.; Melo, A.; Speck-Planche, A.; Cordeiro, M. N. Computer-aided nanotoxicology: assessing cytotoxicity of nanoparticles under diverse experimental conditions by using a novel QSTR-perturbation approach *Nanoscale* 2014, 6, 10623–30 DOI: 10.1039/C4NR01285B
- (34) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Ds Cordeiro, M. N. Computational modeling in nanomedicine: prediction of multiple antibacterial profiles of nanoparticles using a quantitative structure-activity relationship perturbation model *Nanomedicine (London, U. K.)* 2015, 10, 193–204 DOI: 10.2217/nnm.14.96
- (35) Toropov, A. A.; Toropova, A. P. Quasi-QSAR for mutagenic potential of multiwalled carbon-nanotubes *Chemosphere* 2015, 124, 40–46 DOI: 10.1016/j.chemosphere.2014.10.067
- (36) Toropov, A. A.; Leszczynska, D.; Leszczynski, J. Predicting water solubility and octanol water partition coefficient for carbon nanotubes based on the chiral vector *Comput. Biol. Chem.* 2007, 31, 127–8 DOI: 10.1016/j.compbiolchem.2007.02.002
- (37) Toropova, A. P.; Toropov, A. A.; Rallo, R.; Leszczynska, D.; Leszczynski, J. Nano-QSAR: genotoxicity of multi-walled carbon nanotubes *Int. J. Environ. Res.* 2016, 10, 59–64
- (38) González-Durruthy, M.; Monserrat, J. M.; Alberici, L. C.; Naal, Z.; Curti, C.; González-Díaz, H. Mitoprotective activity of oxidized carbon nanotubes against mitochondrial swelling induced in multiple experimental conditions and predictions with new expected-value perturbation theory *RSC Adv.* 2015, 5, 103229–103245 DOI: 10.1039/C5RA14435C
- (39) Poincaré, H. Sur le problème des trois corps et les équations de la dynamique *Ac. Math.* 1890, 13, 1–270
- (40) Bouzarth, E. L.; Brooks, A.; Camassa, R.; Jing, H.; Leiterman, T. J.; McLaughlin, R. M. Epicyclic orbits in a viscous fluid about a precessing rod: theory and experiments at the micro- and macro-scales *Phys. Rev. E* 2007, 76, 016313 DOI: 10.1103/PhysRevE.76.016313
- (41) Urabe, H.; Sugawara, Y.; Ataka, M.; Rupprecht, A. Low-frequency Raman spectra of lysozyme crystals and oriented DNA films: dynamics of crystal water *Biophys. J.* 1998, 74, 1533–40 DOI: 10.1016/S0006-3495(98)77865-8
- (42) Chou, K. C. Identification of low-frequency modes in protein molecules *Biochem. J.* 1983, 215, 465–9 DOI: 10.1042/bj2150465
- (43) Chou, K. C. The biological functions of low-frequency vibrations (phonons). VI. A possible dynamic mechanism of allosteric transition in antibody molecules *Biopolymers* 1987, 26, 285–95 DOI: 10.1002/bip.360260209
- (44) Hennrich, F.; Krupke, R.; Lebedkin, S.; Arnold, K.; Fischer, R.; Resasco, D. E.; Kappes, M. M. Raman spectroscopy of individual single-walled carbon nanotubes from various sources *J. Phys. Chem. B* 2005, 109, 10567–73 DOI: 10.1021/jp0441745
- (45) Fernandez-Lozano, C.; Cuinas, R. F.; Seoane, J. A.; Fernandez-Blanco, E.; Dorado, J.; Munteanu, C. R. Classification of signaling proteins based on molecular star graph descriptors using Machine Learning models *J. Theor. Biol.* 2015, 384, 50–8 DOI: 10.1016/j.jtbi.2015.07.038
- (46) Fernandez-Lozano, C.; Gestal, M.; Pedreira-Souto, N.; Postelnicu, L.; Dorado, J.; Munteanu, C. R. Kernel-based feature selection techniques for transport proteins based on star graph topological indices *Curr. Top. Med. Chem.* 2013, 13, 1681–91 DOI: 10.2174/15680266113139990119
- (47) Fernandez-Blanco, E.; Aguiar-Pulido, V.; Munteanu, C. R.; Dorado, J. Random Forest classification based on star graph topological indices for antioxidant proteins *J. Theor. Biol.* 2013, 317, 331–7 DOI: 10.1016/j.jtbi.2012.10.006
- (48) Perez-Bello, A.; Munteanu, C. R.; Ubeira, F. M.; De Magalhaes, A. L.; Uriarte, E.; Gonzalez-Diaz, H. Alignment-free prediction of mycobacterial DNA promoters based on pseudo-folding lattice network or star-graph topological indices *J. Theor. Biol.* 2009, 256, 458–66 DOI: 10.1016/j.jtbi.2008.09.035
- (49) Vázquez, J. M.; Aguiar, V.; Seoane, J. A.; Freire, A.; Serantes, J. A.; Dorado, J.; Pazos, A.; Munteanu, C. R. Star Graphs of Protein Sequences and Proteome Mass Spectra in Cancer Prediction *Curr. Proteomics* 2009, 6, 275–288 DOI: 10.2174/157016409789973752
- (50) Fernandez-Blanco, E.; Rivero, D.; Rabunal, J.; Dorado, J.; Pazos, A.; Munteanu, C. R. Automatic seizure detection based on star graph topological indices *J. Neurosci. Methods* 2012, 209, 410–9 DOI: 10.1016/j.jneumeth.2012.07.004
- (51) Pedersen, P. L.; Greenawalt, J. W.; Reynafarje, B.; Hullihen, J.; Decker, G. L.; Soper, J. W.; Bustamante, E. Preparation and characterization of mitochondria and submitochondrial particles of rat liver and liver-derived tissues *Methods Cell Biol.* 1978, 20, 411–81 DOI: 10.1016/S0091-679X(08)62030-0

- (52) Munteanu, C. R.; Magalhaes, A. L.; Duardo-Sanchez, A.; Pazos, A.; Gonzalez-Diaz, H. S2SNet: A Tool for Transforming Characters and Numeric Sequences into Star Network Topological Indices in Chemoinformatics, Bioinformatics, Biomedical, and Social-Legal Sciences *Curr. Bioinf.* 2013, 8, 429– 437 DOI: 10.2174/1574893611308040005
- (53) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.
- (54) Randić, M.; Zupan, J.; Vikić-Topić, D. On representation of proteins by star-like graphs *J. Mol. Graphics Modell.* 2007, 26, 290– 305 DOI: 10.1016/j.jmngm.2006.12.006
- (55) Munteanu, C. R. CNT Raman for Mitochondrial Oxygen Consumption Data Set and Regression Models. <https://dx.doi.org/10.6084/m9.figshare.3472349> (accessed Jan 31, 2017).
- (56) Wold, S.; Ruhe, A.; Wold, H.; Dunn, W. J., III The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses *SIAM J. Sci. Stat Comput* 1984, 5, 735– 743 DOI: 10.1137/0905052
- (57) Bishop, C. M. *Neural Networks for Pattern Recognition*; Oxford University Press: USA, 1995.
- (58) Breiman, L. *Random Forests Machine Learning* 2001, 45, 5– 32 DOI: 10.1023/A:1010933404324
- (59) Lanza, I. R.; Nair, K. S. Functional assessment of isolated mitochondria in vitro *Methods Enzymol.* 2009, 457, 349– 72 DOI: 10.1016/S0076-6879(09)05020-4
- (60) Galano, A.; Francisco-Marquez, M.; Martínez, A. Influence of Point Defects on the Free-Radical Scavenging Capability of Single-Walled Carbon Nanotubes *J. Phys. Chem. C* 2010, 114, 8302– 8308 DOI: 10.1021/jp101544u
- (61) Francisco-Marquez, M.; Galano, A.; Martínez, A. On the Free Radical Scavenging Capability of Carboxylated Single-Walled Carbon Nanotubes *J. Phys. Chem. C* 2010, 114, 6363– 6370 DOI: 10.1021/jp100065t
- (62) Galano, A. Influence of Diameter, Length, and Chirality of Single-Walled Carbon Nanotubes on Their Free Radical Scavenging Capability *J. Phys. Chem. C* 2009, 113, 18487– 18491 DOI: 10.1021/jp904646q
- (63) Martínez, A.; Galano, A. Free Radical Scavenging Activity of Ultrashort Single-Walled Carbon Nanotubes with Different Structures through Electron Transfer Reactions *J. Phys. Chem. C* 2010, 114, 8184– 8191 DOI: 10.1021/jp100168q
- (64) Martínez, A.; Francisco-Marquez, M.; Galano, A. Effect of Different Functional Groups on the Free Radical Scavenging Capability of Single-Walled Carbon Nanotubes *J. Phys. Chem. C* 2010, 114, 14734– 14739 DOI: 10.1021/jp1033382
- (65) Geng, J.; Kim, K.; Zhang, J.; Escalada, A.; Tunuguntla, R.; Comolli, L. R.; Allen, F. I.; Shnyrova, A. V.; Cho, K. R.; Munoz, D. Stochastic transport through carbon nanotubes in lipid bilayers and live cell membranes *Nature* 2014, 514, 612– 5 DOI: 10.1038/nature13817
- (66) Tunuguntla, R. H.; Allen, F. I.; Kim, K.; Belliveau, A.; Noy, A. Ultrafast proton transport in sub-1-nm diameter carbon nanotube porins *Nat. Nanotechnol.* 2016, 11, 639– 44 DOI: 10.1038/nnano.2016.43