**Journal of Nanotoxicology and Nanomedicine**
Volume 2 • Issue 1 • January-June 2017

# Exploring Simple, Interpretable, and Predictive QSPR Model of Fullerene $C_{60}$ Solubility in Organic Solvents

Lyudvig S. Petrosyan, Department of Physics, Jackson State University, Jackson, MS, USA

Supratik Kar, Interdisciplinary Center for Nanotoxicity, Department of Chemistry and Biochemistry, Jackson State University, Jackson, MS, USA

Jerzy Leszczynski, Interdisciplinary Center for Nanotoxicity, Department of Chemistry and Biochemistry, Jackson State University, Jackson, MS, USA

Bakhtiyor Rasulev, Department of Coatings and Polymeric Materials, North Dakota State University, Fargo, ND, USA

## ABSTRACT

Buckminsterfullerene ($C_{60}$) and its derivatives have currently been used as promising nanomaterial for diagnostic and therapeutic agents. They are applied in pharmaceutical industry due to their nanostructure characteristics, stability and hydrophobic character. Due to its sparingly soluble nature, the solubility of $C_{60}$ has been of enormous attention among carbon nanostructure investigators owing to its fundamental importance and practical interest in nanotechnology and medical industry. In order to study the diverse role of $C_{60}$ and its derivatives the dependence of fullerene's solubility on molecular structure of the solvent must be understood. Current study was dedicated to the exploration of the solubility of fullerene $C_{60}$ in 156 organic solvents using simple, interpretable and predictive 1D and 2D descriptors employing quantitative structure-property relationship (QSPR) technique. The authors employed genetic algorithm followed by multiple linear regression analysis (GA-MLR) to generate the correlation models. The best performance is accomplished by the four-variable MLR model with internal and external prediction coefficient of $Q^2 = 0.86$ and $R^2_{pred} = 0.89$. The study identified vital properties and structural fragments, particularly valuable for guiding future synthetic as well as prediction efforts. The model generated with the highest number of organic solvents to date with simple descriptors can be reproduced in no time to predict the solubility of $C_{60}$ in any new or existing organic solvents. This approach can be used as an efficient predictor for fullerenes' solubility in various organic solvents.

## KEYWORDS

$C_{60}$, Chemometrics, Fullerene, GA, MLR, QSPR, Solubility

## 1. INTRODUCTION

Fullerene, a highly symmetrical cage-like molecule has specific interaction with organic solvents and its knowledge can provide significant information on the mechanisms of solute-solvent interactions. The fullerenes have defined rigid geometries in distinction to other solutes whose shapes undergo conformational changes. Not only that intramolecular vibrational partition functions may undergo solvent-dependent changes (Prylutskyy et al., 2003). Due to sparingly soluble nature of $C_{60}$ in major organic solvents, the production cost is still high for this nanomaterial (Shunaev et al., 2015). Therefore, understanding of fullerene's solubility provides significant feature assisting in purification,

DOI: 10.4018/JNN.2017010103

extraction, bioavailability, reactivity, and risk assessment of fullerenes. This information is vital due to ample application of carbon nanostructures, such as $C_{60}$ and its derivatives in diverse aspects of nanotechnology, pharmaceuticals, cosmetic, medicinal chemistry, environmental applications (Cook et al., 2010; Bogatu, & Leszczynska, 2016) and materials science (Gharagheizi & Alamdari, 2008; Sivaraman et al., 2001).

Quantitative structure-property relationship (QSPR) represents a powerful tool for modeling and prediction of physiochemical properties. The QSPR method is defined on the foundation of a mathematical algorithm providing a rational basis for establishing a predictive correlation model. Apart from providing a mathematical correlation, it also enables the exploration of chemical features encoded within parameters (descriptors) (Roy, Kar, & Das, 2015a; Toropova, 2016). Hence, diverse set of descriptors plays a noteworthy role in the recognition as well as analysis of the chemical basis involved in a studied phenomenon. Therefore, reliable QSPR model can offer time and cost-effective measure of $C_{60}$ solubility values in organic solvents in the absence of experimental data.

A series of investigation for predicting $C_{60}$ solubility in organic solvents employing QSPR model has been reported in the last 12 years. Liu et al. (2005) generated a linear model as well as a least-squares support vector machine (LSSVM) model for predicting the solubility of $C_{60}$ in 128 and 122 organic solvents, respectively. Toropov et al. (2007, 2009) demonstrated two kinds of descriptors methods for predicting solubility of $C_{60}$ in different organic solvents. Same dataset was used to build one-variable model once with the optimal descriptors calculated with simplified molecular input line entry system(SMILES) (Toropov et al., 2007) and in another work with International Chemical Identifier (InChI) (Toropov et al., 2009) with high statistical results. Petrova et al. (2011) depicted successful application of the GA-MLR technique in combination with quantum-chemical and topological descriptors yields reliable four-variable models for 122 organic solvents. One GA-MLR model was developed to predict the fullerene solubility in 36 benzene derivatives by Pourbasheer et al. (2011). Ghasemi et al. (2013) proposed first 3D-QSAR model employing VolSurf based descriptors with SPA-SVM (successive projection algorithm-support vector machine) method to predict $C_{60}$ solubility in 132 organic solvents with acceptable statistical results. In recent time, Xu et al. (2016) proposed a QSPR model for predicting the solubility of fullerene $C_{60}$ in 156 diverse organic solvents with the norm indexes.

In this regard, we aimed to find simple, predictive, computationally time-efficient and mechanistically interpretable model to predict the solubility of $C_{60}$ in the same set of organic solvents considered by Xu et al. (2016). In addition, the study intends to estimate predictive potential of the simple 1D and 2D descriptors to model the solubility of the fullerene $C_{60}$ in a large number of organic solvents.

## 2. MATERIALS AND METHODS

### 2.1. Data Set

The experimental solubility (S) data of $C_{60}$ in 156 organic solvents (Table 1) were collected from two datasets: Beck and Mándi (1997) and Semenov et al. (2010). As the logarithmic values of molar fractions corresponded to the free energy changes in the solvation process, the unit of solubility was considered as logS, instead of weight units (e.g., mg/mL).