

# Solvent molecular descriptors on poly(D, L-lactide-co-glycolide) particle size in emulsification–diffusion process

Zelimir Jelcic

PLIVA-Research Institute Ltd., Pharmaceutical Technology, HR10000 Zagreb, Prilaz Baruna Filipovica 25, Zagreb Croatia

Received 4 July 2003; accepted 29 March 2004

Available online 19 June 2004

## Abstract

The size of poly(D, L-lactide-co-glycolide) (PLGA) nanoparticle was influenced by the thermodynamic parameters (after [Sung-Wook Choi, Hye-Young Kwon, Woo-Sik Kim, Jung-Hyun Kim, Colloids Surf. A Physicochem. Eng. Aspects 201 (2002) 283–289.]) such as mutual diffusion coefficients (exchange ratio,  $R$ ) and solvent–polymer interaction parameters ( $\chi$ ). However, this study introduces the concept of solvent molecular descriptors for elucidating the effect of solvent on the PLGA nanoparticle size in the emulsification–diffusion method. The most pronounced effects on the nanoparticle sizes have the empirical hydrophilicity index  $H_y$ , and the molecular descriptors such as the Randic shape descriptor. More rigor results that various molecular descriptors have on the PLGA nanoparticle size are obtained by the molecular descriptors that correspond to the 3D structure of molecules (3D MoRSE descriptors, directional WHIM indices and the GETAWAY [Geometry, Topology and Atom-Weights Assembly] molecular descriptors). The derived experimental data correlated well with information available from other procedures, confirming that this new approach provides complementary insight into the thermodynamics and the molecular nature of the polymer nanoparticle formation in the solvent type of chemical environments.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Emulsification–diffusion method; Poly(D, L-lactide-co-glycolide) nanoparticle; Diffusion coefficient; Exchange ratio; Solvent–polymer interaction parameter; Molecular descriptors

## 1. Introduction

The formation of polymer nanoparticles is of interest in the pharmaceutical industry to increase the bioavailability of drugs. Nanoparticles from preformed polymers based on emulsification–diffusion technique are prepared by emulsifying an organic solution of the drug in an aqueous solution of a stabilising agent followed by rapid displacement of the solvent from the internal into the external phase, provoking solid particle formation [1]. Complex hydrodynamic processes at the interfacial area lead to the creation of the nanoparticles [2,3]. The formation of nanoparticles at high solids loading of materials that are sparingly soluble requires rapid micromixing so that nucleation and growth occurs under uniform conditions. Furthermore a mechanism for arresting the growth nanoparticles must be devised to prevent aggregation and growth past the desired size [4]. In general, the selections of solvent and preparation

conditions are the most important parameters affecting the characteristics of polymeric nanospheres. It is believed that the nanospheres formation involves the mechanism as described [5]: stirring causes the dispersion of the solvent as irregular sized globules in equilibrium with the continuous phase and the stabilizer is then absorbed on the larger interface created; homogenization further results in the smaller globules; the addition of water destabilizes the equilibrium and the diffusion of the organic solvent to the external surface causes local supersaturation near the interface, and nanoparticles are formed, due to the phase transformation and polymer aggregation that occur in these regions [6]. During the transport of solute, globules are produced in a final colloidal suspension free of organic solvent. The emulsification–diffusion method has been used successfully to prepare biodegradable nanoparticles in an efficient and reproducible manner. The hypothesis of thermodynamic stabilization of the nanoparticles of a certain size of the particles by the surfactant molecules has been proposed [7].

Choi et al. [8] have recently quantitatively studied the emulsification–diffusion method for making poly(D,

E-mail address: zelimir.jelcic@pliva.hr (Z. Jelcic).

L-lactide-co-glycolide) (PLGA) nanoparticles considering the thermodynamic parameters, such as diffusion coefficients ( $D_{sw}$ ,  $D_{ws}$ ), exchange ratio ( $R = D_{sw}/D_{ws}$ ), and solvent–polymer interaction parameter ( $\chi$ ). They have found that  $R$  was proportional to the size of the PLGA nanoparticles, while  $\chi$  was inversely proportional to it. In the case of the higher value of  $R$  and lower value of  $\chi$ , a small local supersaturation region was produced at the O/W interface and the small nanoparticles separated from the oil globule were formed in that region. The partially water-soluble solvents were used for the preparation of nanoparticles: ethyl acetate (EtAc), methyl ethyl ketone (MEK), propylene carbonate (PC), and benzyl alcohol (BA). Choi et al. [8] explained the mechanism of the emulsification–diffusion method as a combination of solvent–water exchange effect and polymer–solvent interaction effect. This simple estimation method is based on correlation with experimental physicochemical quantities, such as diffusion coefficients ( $D$ ) and solvent solubility ( $\delta$ ). The operational/process parameters were kept constant in the Choi et al. [8] studies, allowing us to develop a new approach for judging the effect of solvent on the emulsification–diffusion process by correlating nanoparticle size with the parameters being calculated only from molecular structure, such as molecular descriptors.

The use of molecular descriptors to predict quantitatively environmental or pharmacological activities or properties such as boiling points, vapour pressures, or chromatographic retention indices of a homologous series of organic molecules is a modern topic in computational chemistry. Several physicochemical properties of a compound can be derived from its structure. Among the properties that can be computed are for instance the octanol/water partition coefficient  $\log P$ ,  $pK_a$ , the molecular polarizability and the standard heat of formation. Since all of these properties can be computed also for different solvents, identification of desirable solvents can be done prior to nano particle fabrication [9]. In this paper, the size of polymer nanoparticles prepared by emulsification–diffusion process can be described with only a single parameter, for example by the molecular descriptor hydrophilicity factor,  $H_y$ .

## 2. Methods

### 2.1. Computational methods

Solvents molecular descriptors for modeling were computed using the package DRAGON of Todeschini and Consonni [13]. These descriptors contain information about the whole molecule in terms of size, shape, symmetry and atom distribution. These indices are calculated from the cartesian  $x$ ;  $y$ ;  $z$  coordinates of a molecule within different weighting schemes in a straightforward manner. The input files for descriptors calculation were obtained from databases and checked by the structural imaging (ViwerLite 5.0, Accelrys Inc.) and the geometry optimization using

the semiempirical method AM1 [10] implemented in the ArgusLab software [11]. A set of molecular descriptors was determined for solvents: (a) 1D-descriptors: counting of atoms and fragments, MW and sum of atomic properties; (b) 2D-descriptors: topological (from molecular graph); (c) 3D-descriptors WHIM (Weighted Holistic Invariant Molecular) descriptors [12] GETAWAY [Geometry, Topology and Atom-Weights Assembly], that contain information about the whole 3D-molecular structure in terms of size, symmetry and atom distribution. The solvent WHIM 3D-molecular descriptors are calculated from a weighted covariance matrix obtained by the molecule geometrical coordinates [13]. These descriptors contain information concerning size, symmetry, shape and distribution of the molecule atoms.

The correlation of 420 descriptors (having different values for this solvent set) with a given molecular property was calculated (Statgraphics for Windows, version 2.1, Statistical Graphics Corp.) and those with small or no correlation (smaller than a given cut-off) to nanoparticle size were discarded. Overall, 133 descriptors (out of 420) were found to be statically significant at the  $P$ -level of 0.05. The final model obtained with single molecular descriptors satisfies the consideration that for a study of structure–property relationship it should be obtained correlation coefficients above 0.95. The most predictive molecular descriptors were identified in several stages. At first, the statistical quality of a single-descriptor logistic model, the  $P$ -value, was assessed for each of the descriptors. Descriptors with the  $P$ -value above 0.05 were then omitted from further analysis. The remaining potentially predictive descriptors were subsequently analyzed. All possible combinations of 1, 2, 3 and 4-descriptor models, along with a stepwise regression algorithm, which does not restrict the number of descriptors in the model cannot be persuaded due to small number of response data points. Some of the couples of molecular descriptor indices are perfectly correlated (with the coefficient 1), due to small numbers of solvents. It indicates that the two indices of each pair contain the same information. Among those descriptors highly correlated to themselves, the ones that can be most easily interpreted were selected. The data set was reduced to several descriptors, which are most important, and their respective correlations with the polymer nanoparticle size were analyzed.

### 2.2. Thermodynamic parameters

The (average) polymer particle size is determined by the solvent–polymer interaction parameter (Fig. 4, ref. [8]). The particle size decreases non-linearly with increase in the solvent–polymer interaction parameter,  $\chi$ . The particle size can be described by power function:

$$d = 122.16(\chi)^{-0.671} \quad (1)$$

where  $d$  is the average particle size (in nm) and  $\chi$  is the polymer–solvent interaction parameter, with  $R^2 = 0.897$ .

The particle size cannot be described by a linear function of solvent–polymer interaction parameter  $\chi$ .

The average particle size, also, depends on the mutual diffusion coefficients ratio, the exchange ratio  $R$  (Fig. 3, ref. [8]). The particle size non-linearly increases with the increase in the exchange ratio  $R$  ( $D_{sw}/D_{ws}$ ). The particle size can be described by exponential function:

$$d = 109.72 \exp(0.2381R) \quad (2)$$

where  $d$  is the average particle size (in nm) and  $R$  is the mutual diffusion coefficients ratio, with  $R^2 = 0.9376$ . The linear function is also not appropriate for description of particle size as a function of the exchange ratio  $R$ .

### 2.3. Solvent molecular descriptors

The relationships between the molecular structures of organic compounds and their chemical or properties can be investigated using quantitative structure–property relationship (QSPR) methods [14]. In the simplest case, this approach involves linear relationships between an experimentally measured property and some molecular descriptor or substituent constant scale, derived either from another experiment or calculated based on the chemical or electronic structure of the compounds [15]. However, in many cases single-parameter linear equations are of unacceptably low statistical quality and predictive power that can result in a subsequent dramatic loss of the physical meaning of the correlations obtained [16]. General classes of descriptors include topological, geometrical, electronic, and hybrid representations of the molecules. Topological descriptors are calculated directly from the connection table representation of the structure and employ methods drawn from mathematical graph theory. Geometric descriptors are calculated from 3D molecular models, which are generated with molecular orbital methods. Electronic descriptors come from empirical or molecular orbital calculations. The descriptor selection can be applied and correlated, using statistical methods, in principle, to any chemical property that is determined by the molecular structure, and it has been applied to a wide variety of chemical properties. This structure–property relationship can model many types of solute/solvent interaction-dependent chemical properties. These properties are related to molecular structure through component contributions to the energy required to stabilize a solvent–molecule envelope around a solute and the energies gained/lost from solute–medium electrostatic and hydrogen bonds. The molecular descriptor is used to find quantitative mathematical relationships between the intrinsic molecular structure and observable properties of chemical compounds, based on a tacit assumption, which can be expressed in terms of the following equation:

$$\begin{aligned} \text{Property} &= F(\text{molecular structure}) \\ &= F(\text{molecular descriptor}) \end{aligned} \quad (3)$$

where the property is the magnitude of measured effect produced by a molecule under preset processing conditions and right-hand of the equation represents either an empirical property or a theoretical parameter for the total molecular structure or relevant substructural fragment(s) [17]. This approach leads to the belief that a proper choice of molecular descriptors will give a precise prediction of properties for molecules. The topological, geometrical, and quantum chemical indices, the three groups of theoretical parameters, which associates a real number with the chemical structure based on certain aspects of chemical structure, are numerical descriptors of molecular architecture and quantify such aspects of molecular structure as size, shape, symmetry, complexity, branching, etc. An advantage of the exclusive use of the theoretical descriptors is that they are free of the uncertainty of experimental measurements. Andersson et al. [18] have compared the information content in measured physicochemical and some calculated descriptors. Their results show that the descriptor sets contain similar information for the data sets investigated. Thus, the work presented is focused on the use of calculated descriptors, as distinction to the Choi et al. [8] approach that used only measured (or derived) properties.

The partitioning of polymer into different solvent/non-solvent phase compartments depends mainly on the physico-chemical properties of the studied chemicals. The effect of the interaction between solvent and water should be incorporated in the model, as this is reflected in different particle size depending on the quality of solvent [19]. However, due to strong non-linear particle size dependence on the thermodynamic parameters [8], another approach should be tried. The influence that the solvent properties descriptors have on the nanoparticle size prepared in the preset processing “window”, is determined here by more rigor molecular descriptor parameters.

### 2.4. Correlation of molecular descriptors and nanoparticle size

Use of a more water–miscible solvent has been shown to improve solvent extraction for preparing polymeric spheres [20]. Thus, the empirical molecular descriptor, hydrophilic factor  $H_y$ , is the first choice for describing such tendency. The  $H_y$  describes hydrophilicity, the tendency of a solvent molecule to be solvated by water. Hydrophilicity index ( $H_y$ ) is a simple empirical index based on the counts descriptors obtained as a function of the count of the hydrophilic groups and it is defined as [13]:

$$H_y = \frac{(1 + N_{Hy}) \log_2(1 + N_{Hy}) + N_C((1/A) \log_2(1/A)) + \sqrt{N_{Hy}/A^2}}{\log_2(1 + A)} \quad (4)$$

where  $N_{Hy}$  is the number of hydrophilic groups (-OH, -NH, -SH),  $N_C$  the number of carbon atoms and  $A$  the number of atoms (hydrogen excluded).

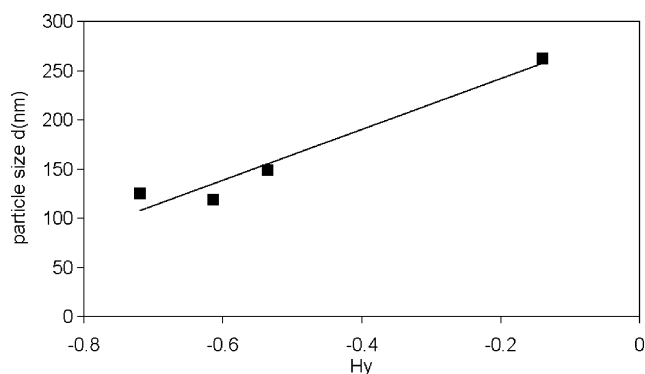


Fig. 1. The correlation of PLGA particle size  $d$  (nm) with the empirical hydrophilic factor  $Hy$ ; full line:  $d$  (nm) = 258.4( $Hy$ ) + 293.43;  $R^2$  = 0.9535.

The particle size in emulsion–diffusion process [8] show a good linear correlation with the empirical solvent hydrophilic molecular descriptor  $Hy$  (Fig. 1). In the range of nanoparticles produced by emulsification–diffusion process the solvent hydrophilic descriptors  $Hy$  is highly correlated with the nanoparticle size ( $R^2 = 0.9535$ ); however, the comparable linear correlation with the interaction parameter  $\chi$  has much lower correlation factor of  $R^2 = 0.7176$ . Construction of linear regression models containing non-linear terms is most often prompted when the data is clearly not well fitted by a linear model, but where regularity in the data suggests that some other model will fit. Non-linear dependency of properties on molecular or/and topological parameters became apparent early in the development of models and a first approach to the solution of these problems involved fitting a non-linear function [21]. The corresponding power function of the particle size versus. the hydrophilic descriptor  $Hy$  is given by:

$$d = 103.63(-Hy)^{-0.4735} \quad (5)$$

where  $d$  is the average particle size (in nm) and  $(-Hy)$  is the negative value of the hydrophilic molecular descriptor, with  $R^2 = 0.9622$ . However, such non-linear function does not improve the model.

We can conclude that the combination of solvent/water exchange effect and polymer/solvent interaction effect, that Choi et al. [8] used to explain the mechanism for making polymer particles, can be simulated, even better, with single parameter, and that is the solvent hydrophilic molecular descriptor  $Hy$ .

The hydrophilic descriptor  $Hy$  is positively correlated with the GETAWAY molecular descriptor  $H2p$ , a H autocorrelation of lag 2 (weighted by atomic polarizabilities) ( $R = 0.9989$ ), and with the 3D MoRSE molecular descriptors (calculated by summing atom weight by a different angular scattering function) [22]:  $Mor14u$  ( $R = 0.9988$ ),  $Mor14e$  ( $R = 0.9984$ ),  $Mor16u$  ( $R = -0.9969$ ), where the 3D MoRSE descriptors are:  $Mor14u$  signal 14 unweighted,  $Mor14e$  signal 14 weighted by atomic Sandersons electronegativities, and  $Mor16u$  signal 16 unweighted. The hydrophilic

descriptor  $Hy$  is positively correlated with the topological descriptor, the polarity number  $Pol$  ( $R = 0.9980$ ). Another set of good, positive correlations is found with the GETAWAY molecular descriptors (calculated from the elements of the leverage matrix obtained by the centred atomic coordinates (molecular influence matrix);  $R+$  descriptors are obtained from the leverage/geometry matrix) of  $R$  maximal autocorrelation of lag five weighted by atomic van der Waals volumes ( $R5v+$ ,  $R = 0.9984$ ), weighted by atomic masses ( $R5m+$ ,  $R = 0.9974$ ) and weighted by atomic polarizabilities ( $R5p+$ ,  $R = 0.9973$ ).

The hydrophilic descriptor  $Hy$  is negatively correlated ( $R = -0.9900$ ) with the Randic molecular profile [23],  $SHP2$ , the average shape profile index of order two (a molecular descriptor derived from the distance distribution moment of the geometry matrix, defined as the average row sum of its entries raised at the second power, normalized by the factor 2!). The hydrophilic descriptor  $Hy$  is negatively correlated with the 3D MoRSE molecular descriptors (obtained by radial functions centred on different interatomic distances):  $Mor13m$  ( $R = -0.9939$ ) and  $Mor13e$  ( $R = -0.9870$ ), where are the 3D MoRSE descriptors:  $Mor13m$ , signal 13 weighted by atomic masses, and  $Mor13e$ , signal 13 weighted by atomic Sandersons electronegativities. Also, negative correlations are observed with the third component accessibility directional WHIM indicies (obtained as statistical indicies of the atoms projected onto the three principal components obtained from weighted covariance matrices of the atomic coordinates): unweighted  $E3u$  ( $R = -0.9895$ ), and  $E3p$  weighted by atomic polarizabilities ( $R = -0.9864$ ). Slightly weaker correlation has been found for the third component size directional WHIM indicies:  $L3u$  unweighted ( $R = -0.9894$ ),  $L3p$  weighted by atomic polarizabilities ( $R = -0.9877$ ), and  $L3v$  weighted by van der Waals volumes ( $R = -0.9862$ ).

These correlations with the empirical molecular descriptor hydrophilicity  $Hy$  indicate that these corresponding molecular descriptors could be also suitable for the modeling of the nanoparticle size in emulsion–diffusion process [8]. The nanoparticle size function may be described by any of here given molecular descriptors (positively or negatively correlated with hydrophilic factor  $Hy$ ).

It is interesting that the best correlation with the nanoparticle size shows a topological index,  $ATS3v$  (Fig. 2), which is essentially a numerical molecular descriptor, encoding both molecular structure and physico-chemical properties of a molecule [24]. Moreau and Broto autocorrelation vectors are calculated from the molecular graph, by summing the products of atom weights of the terminal atoms of all paths of the considered path length (the lag) [25]. Molecular surface properties such as the electrostatic or the hydrophobicity potential are condensed into an autocorrelation descriptor [26]. This 2D autocorrelation descriptor has been used to estimate log  $P$ -values [27].

The second best non-empirical solvent molecular descriptor is the Randic molecular profile [28]  $SHP2$ , the average



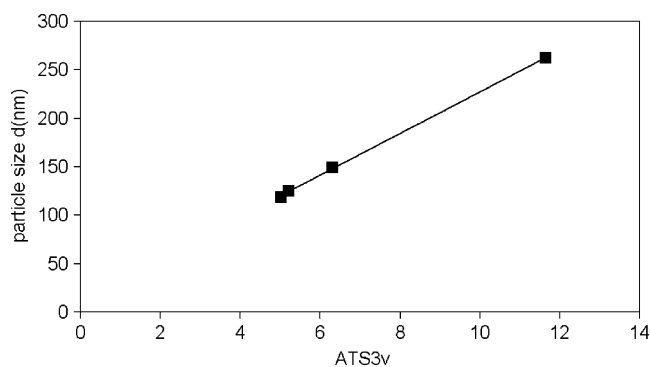


Fig. 2. The particle size correlation with the 2D molecular descriptor ATS3v, Broto-Moreau autocorrelation of a topological structure-lag three weighted by atomic van der Waals volumes; full line:  $d$  (nm) =  $21.455(\text{ATS3v}) + 12.672$ ;  $R^2 = 0.9995$ .

shape profile index of order second, giving the particle size by (almost perfect) linear function (Fig. 3). The correlations of molecular descriptors and nanoparticle size  $d$  (nm) are even better than that with the empirical hydrophilic factor Hy. The Randic index can be interpreted as the relative molecular area of the external accessibility [29]. This area represents the total area accessible from the environment surrounding the molecule. From the constitutional descriptors, a good correlation ( $R^2 = 0.9959$ ) with particle size shows the mean atomic polarizability (scaled to carbon atom) (Mp) (Fig. 4). Atomic polarizability is a sum over all atoms in the molecule and describing the molecule's ability to polarize in a magnetic field, the more polarizable molecules are more toxic. This constitutional descriptor reflects the importance of the bulk of the (solvent) compound and the hydrophobically active surface area of the molecule in the nanoparticle formation. The good correlation ( $R^2 = 0.9949$ ) is observed also for the topological descriptor, the conventional bond-order ID number (piID). BCUT descriptors (based on molecular graphs as a source of different probability distributions), and obtained from the positive and negative eigenvalues of the adjacency matrix, weighting the diagonal elements with atomic weights, are fairly good correlated with the particle size, especially the BELm3,

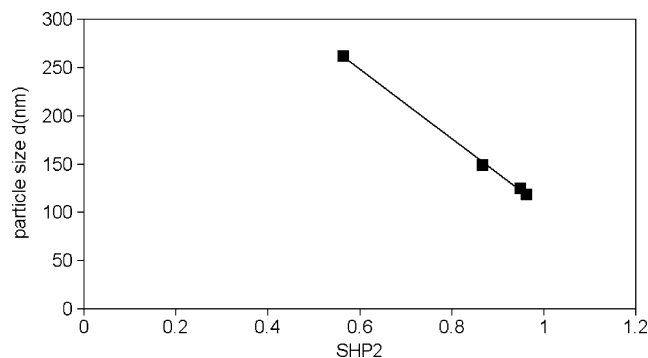


Fig. 3. The particle size correlation with the Randic molecular profile SHP2, the average shape profile index of order 2; full line:  $d$  (nm) =  $-359.92(\text{SHP2}) + 464.33$ ;  $R^2 = 0.9987$ .

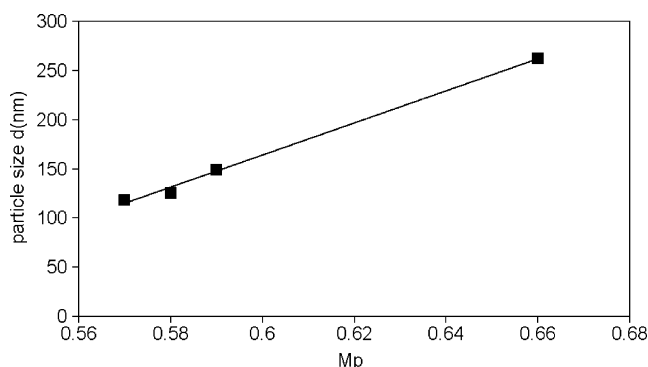


Fig. 4. The particle size correlation with the mean atomic polarizability (scaled to carbon atom) (Mp); full line:  $d$  (nm) =  $1635.5(\text{Mp}) - 817.58$ ;  $R^2 = 0.9959$ .

lowest eigenvalue number three of burden matrix weighted by atomic masses ( $R^2 = 0.996$ ). Burden [30] has developed a simple approach to convert the molecular structure into a matrix and its both lowest eigenvalues uniquely describe the molecule. GETAWAY group of descriptors [31,32] which are based on a leverage matrix similar to that defined in statistics and usually used for regression diagnostics. These molecular descriptors try to match the 3D molecular geometry provided by the molecular influence matrix and atom relatedness by molecular topology, with chemical information by using different atomic weight schemes. The most pronounced correlation ( $R^2 = 0.9944$ ) of all GETAWAY descriptors shows the R4p+, the R maximal autocorrelation of lag four weighted by atomic polarizabilities. The geometrical descriptor [33,34] G1, the gravitational index is well correlated ( $R^2 = 0.9948$ ) to the nanoparticle size  $d$  (nm). From the 3D-MoRSE descriptors [35], calculated by summing atom weights viewed by a different angular scattering function, the best correlated ( $R^2 = 0.995$ ) to the nanoparticle size is the Mor13u, 3D-MoRSE signal unweighted.

The molecular descriptor, which is a very simple measure of the hydrophobic/lyophilic character of a given substance (i.e. the molecular partition coefficient in 1-octanol/water system) [36,37] can be modeled by the Moriguchi octanol/water partition coefficient, MlogP [38,39]. However, this molecular descriptor MlogP does not show pronounced correlation ( $R^2 = 0.7211$ ) to the nanoparticle size.

### 3. Discussion

The properties of solvents affect the nanosphere properties mainly by influencing the removal rate of the solvents [40]. Solid polymeric spheres are formed essentially in two steps, i.e., solvent extraction followed by solvent evaporation, and in both processes the solvent properties play key roles [41]. Changes in physical properties such as lipophilicity ( $\log P$ ) or volatility ( $\log Vp$ ) can affect, for example, whether the solvent moves in aqueous phase and perhaps evaporates to air.

The solubility in water of the solvent determines the extraction rate, while the evaporation rate depends on its boiling point. Generally, faster extraction rate, indicated by higher solubility in water, and faster evaporation rate, indicated by lower boiling point, leads to faster solvent removal and faster nanosphere formation. According to the water solubility and boiling point of the solvents, the extraction rate should decrease in the order from miscible down to less soluble solvent in water, while the evaporation rate of least soluble solvent should be the highest. Choi et al. [8] experiments showed that solvent extraction rate (the faster diffusion of solvent to water,  $D_{sw}$ ) was directly related to the increase in the particle size of the nanospheres formed. This implies that, under these experimental conditions, rapid removal of the solvent resulted in spheres with a loose matrix and large size [42].

The proposed nanoparticle formation mechanisms are in line with solvent/water based interplay. Thus, the correlation of the aqueous solubility of solvents with the parameters being calculated only from molecular structure using descriptors of various types [43,44], can be used to interpret the nanoparticle formation. Three types of interactions participate in the solvent/ solute based properties: cavity effects, polarizability and dipolarity, and hydrogen bonding [45]. The molecular volume ( $V$ ) is the key term to aqueous solubility prediction [43]. In order for a solute to enter into aqueous phase, a cavity must be formed for the solute molecule to occupy. Water as a solvent would much prefer to interact with itself than with a nonpolar solute [46], so that the larger molecules have lower aqueous solubility. However, although molecular volume (or  $Mv$ , the mean atomic van der Waals volume, scaled to carbon atom) can by itself predict nanoparticle size to reasonable accuracy ( $R^2 = 0.9775$ ), this relationship is not as well as other descriptor relationships. The electrostatic interaction, that participates in the solvent/solute based properties, the mean atomic polarizability ( $Mp$ , Fig. 4) is the best descriptor for the nanoparticle size. The models of aqueous solubility were derived using only molecular connectivity indices [44]. The first-order ( ${}^1\chi^v$ ) valence molecular connectivity index, is calculated from the non-hydrogen part of the molecule [47]. This topological descriptor, the valence connectivity index [47],  $X1v$ , also shows a fair correlation to nanoparticle size ( $R^2 = 0.9224$ ).

It is very interesting that the similar descriptors are used to describe the critical micelle concentration of surfactants [48,49]. The properties of 77 diverse nonionic surfactants were very well correlated ( $R^2 = 0.983$ ) with only three theoretical molecular fragment descriptors [50]. Significantly, two of these descriptors are topological expressions for the hydrophobic tail and the third descriptor represents the size and polarity of the hydrophilic head. This makes attractive approach for describing the formation of the surfactant-modified nanoparticles.

The boiling point of any chemical can be explained thermodynamically as the temperature at which the thermal energy of the particles is sufficient to break the cohesion forces

which keep the substance in the aggregation state which characterizes the liquid phase, and allows an estimate of the evaporation of chemicals. The boiling point of solvents can be related to the van der Waals forces for non-polar molecules, which are weaker than the dipole forces characteristic for polar molecules [51]. The energy of these kind of intermolecular forces is closely related to structural size and molecular branching [52]. The boiling point [53] or vapour pressure [54] can be modelled by molecular descriptors, and parameters such as molecular weight, volume, surface area and Randic connectivity index are required to model boiling point. Also, the molecular descriptors used to model the boiling point of aliphatic hydrocarbons by the neural networks included the first, second, third, and fourth order valence molecular connectivity indices, a second-order kappa shape index [55,56], dipole moment, and molecular weight [57]. The kappa shape index index quantifies the structure of compounds in terms of its relative starlikeness and straight chainlikeness shape. Of course, the increase in the total area accessible from the environment surrounding the molecule, that is of Randic index, the number of intermolecular interactions increase producing an increment in the boiling point and explaining the correlation of this index with this experimental property. It also explains the success of this index in describing solubility, partition coefficients or nanoparticle size (in emulsion–diffusion method).

Another attempt was made to elucidate mechanisms of formation of nanoparticles in terms of interfacial turbulence between two unequilibrated liquid phases involving flow, diffusion and surface tension decrease [2]. When the polymer solution is dispersed into the aqueous solution, the perturbation of the interface spontaneously produces a larger interfacial area, which leads to nano-sized quasi-emulsion droplets of polymer solution. Long-range forces like electrostatic interactions and van der Waals forces may lead to attraction or repulsion between the different nano droplets as well as between the nano droplets and the interface [58]. This interfacial turbulence would be governed by the well known Marangoni effect [59]. This effect can be attributed to the surface tension gradient induced by the concentration (or temperature) gradient in the interface layer formed in front of the solidifying interface. It is well known that interfacial Marangoni instabilities may dominate the mass transfer across the interface [60], and this mass transfer across the interface induces the formation of nano droplets in the close vicinity of the interface. This disturbance enhancement is suppressed by thermal diffusivity and the dynamic viscosity of the liquid. These physicochemical parameters can be also derived from the structure [61]. However, one needs several molecular descriptors for prediction of these properties [62]. The point that some of them are encoding the size and shape properties of each solvent molecule correspond to the molecular descriptors used here for the correlation with the particle size.

The process of nanoparticle formation depends both on complex processing and on compound physicochemical

properties (including solubility). Therefore, we do not expect that this molecular descriptor model derived using only a small number of diverse solvent compounds will be a highly precise and rugged tool. A much larger data set, presently unavailable in the published literature, would be required to build a model based not only on structural diversity but also on diverse physicochemical properties. However, as shown in figures, this model produces an exact rank ordering. The selection of the of molecular descriptors that are correlated with the nanoparticle size seems to have a natural association with solvent activity, where the 3D structure recognition of a given solvent is responsible for the behaviour in the emulsification–diffusion formation of nanoparticles. However, the sophisticated representation of these descriptor classes impedes a simple interpretation of the mechanism of nanoparticle formation. Thus, we can only rely on these single descriptors models as an instrument of correlating the solvent activity in the nanoparticle formation.

One can argue about difficulties and oversimplifications as well as about advantages and opportunities of molecular descriptor approach to nanoparticle size prediction. Questions have been raised about defining relative activities, about the rational reasons for the observed strong correlation. Notwithstanding the validity of such concerns and the lack of understanding of underlying mechanisms due to the complexity of nanoparticle systems, empirical schemes offer a useful tool to help to filter out from a collection of solvent structures less desirable ones, even if they may not point the most desirable ones [63]. The molecular descriptor approach employed here appears capable of resolving some of the above concerns. The significant difference to Choi et al. [8] approach is the need to compare the same compound (PLGA) with respect to the different solvents. This is a problem in ordering of the solvents in the nanoparticle formation modeling, because in this case, the comparison is performed only between the competing processes of a single polymer. Therefore, even a limited comparison of molecular descriptors with the polymer nanoparticle size model is appropriate.

#### 4. Conclusions

The main goal of the presented study was to evaluate the classes of molecular descriptors that can be used to describe the experimentally observed nanoparticle formation in the emulsification–diffusion process. Our preliminary results demonstrate that the most promising molecular descriptors are derived from either 3D or 2D molecular structure indices, which are based on radial distribution functions, topological indices, or autocorrelation functions. These classes of descriptors seem to be naturally related to the solvent activity as they associate the solvent properties with a 3D structure and shape of the solvent molecules. These results suggest that a molecular descriptor model of nanoparticle formation may be built by using only a single appropriate parameter, although the relevance of the identi-

fied descriptors to the continuous-scale correlation has yet to be shown, due to very limited number of data points.

The data reveal indeed that the particle size correlates almost perfectly (for the given set of experimental conditions) with the hydrophilicity factor,  $H_y$ , and also with other, more dimensional molecular descriptors of the solvent used for the preparation of the starting polymer solution. This is on the one hand an interesting and on the other hand a surprising result as the particle formation mechanism is rather complex involving at least the emulsification step, the solvent re-equilibration after addition of excess water, and finally the build-up of a supersaturation at the droplet interface followed by nucleation and growth. Of course it is clear that the water solubility of the solvents (or the hydrophilic factor) has a strong influence on that kind of particle formation.

The single solvent molecular descriptor can describe the PLGA nanoparticle size in the emulsification–diffusion method. The size of PLGA nanoparticle can be described by empirical molecular hydrophilic descriptor  $H_y$ , but also with the more rigorous molecular descriptors, such as the 3D Randic molecular shape descriptor SHP2.

These results illustrate the tremendous complexity that can be achieved in emulsification–diffusion process and point toward the factors that control such complexity, beyond the thermodynamic parameters [8] such as mutual diffusion coefficients (exchange ratio,  $R$ ) and solvent–polymer interaction parameters ( $\chi$ ). The molecular descriptors approach, in the preset processing “window”, provides a rational basis for the selection of solvent to control the size of PLGA nanoparticles.

From a set of several hundreds molecular descriptors, it was possible to select a few of them to construct linear regression models with good predictability for the final polymer nanoparticle size. We have shown that molecular descriptor method is a quite suitable tool to model nanoparticle size data for a representative set of several solvent compounds encompassing a variety of mechanisms of solvent–polymer interaction parameters and exchange ratio. In fact, fitting equations yield satisfactory predictions for the completely molecular set. We have resorted to correlating and describing properties by linear equations, since the number of molecules within each molecular subset is rather scanty, so that results wouldn't be statistically significant for higher order equations.

Among the molecular descriptors chosen in the present study, the best model of nanoparticle size is that one based on the Randic molecular shape descriptor (i.e. SHP2). Despite different intimate mechanisms occurring between different chemical compounds comprising the present molecular set, we have gotten a highly predictive, global approach.

Present one parameter approach to the prediction of nanoparticle size (under preset processing conditions) appears to have wide potentiality to a number of polymers, at separate processing levels.

Recent results have indeed suggested that the underlying structure/contributions to molecular descriptors are relevant

to nanoparticle formation. Our results suggest that the presented molecular descriptors are useful not only for physical property modeling but also in nanoparticle formation modeling. This study, however, only illustrates the potential of using molecular descriptor methods to aid in the nanoparticle formation development process.

## References

- [1] D. Quintanar-Guerrero, E. Allemann, E. Doelker, H. Fessi, *Pharm. Res.* 15 (1998) 1056–1062.
- [2] H. Fessi, F. Puisieux, J.P. Devissaguet, N. Ammoury, S. Benita, *Int. J. Pharm.* 55 (1989) R1–R4.
- [3] T. Niwa, H. Takeuchi, T. Hino, N. Kunou, Y. Kawashima, *J. Controlled Release* 25 (1993) 89–98.
- [4] T. Trimallie, C. Pichot, A. Elaïssari, H. Fessi, S. Briançon, T. Delair, *Colloid. Polym. Sci.* 281 (12) (2003) 1184–1190.
- [5] M.N.V. Ravi Kumar, U. Bakowsky, C.M. Lehr, <http://www.elsevier.com/gej-ng/10/24/19/49/567/article.pdf> (7.1.2004.).
- [6] D. Quintanar-Guerrero, E. Allemann, E. Doelker, H. Fessi, *Colloid. Polym. Sci.* 275 (1997) 640–647.
- [7] F. Debuigne, J. Cuisenaire, L. Jeunieu, B. Masereel, J.B. Nagy, *J. Colloid Interface. Sci.* 243 (2001) 90–101.
- [8] Sung-Wook Choi, Hye-Young Kwon, Woo-Sik Kim, Jung-Hyun Kim, *Colloids Surf. A Physicochem. Eng. Aspects* 201 (2002) 283–289.
- [9] S. Brochini, *Adv. Drug Deliv. Rev.* 52 (2001) 123–130.
- [10] M.J.S. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, *J. Am. Chem. Soc.* 107 (1985) 3902–3909.
- [11] ArgusLab 3.1.0, Planaria Software, <http://www.arguslab.com>.
- [12] R. Todeschini, P. Gramatica, *Quant. Struct.-Act. Relat.* 16 (1997) 113–119.
- [13] R. Todeschini, V. Consoni, *Handbook of molecular descriptors, methods and principles*, in: R. Mannhold, H. Kubinyi, H. Timmerman (Eds), *Medicinal Chemistry*, vol. 11, Wiley-VCH, Weinheim, Germany, 2000.
- [14] E.M. Duffy, W.L. Jorgensen, *J. Am. Chem. Soc.* 122 (2000) 2878–2888.
- [15] A.R. Katritzky, U. Maran, V.S. Lobanov, M. Karelson, *J. Chem. Inf. Comput. Sci.* 40 (2000) 1–18.
- [16] R. Put, C. Perrin, F. Questier, D. Commas, D.L. Massart, Y. Vander Heyden, *J. Chromatogr. A* 988 (2003) 261–276.
- [17] S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, G.D. Veith, *Math. Model* 8 (1987) 300–305.
- [18] P.M. Andersson, M. Sjöström, S. Wold, T. Lundstedt, *J. Chemom.* 14 (2000) 629–642.
- [19] D. Quintanar-Guerrero, H. Fessi, E. Allemann, E. Doelker, *Int. J. Pharm.* 143 (1996) 133–141.
- [20] R. Bodmeier, J.W. McGinity, *Int. J. Pharm.* 43 (1988) 179–186.
- [21] C. Hansch, *Acc. Chem. Res.* 2 (1969) 232–239.
- [22] <http://zabib.chemie.uni-erlangen.de/external/cic/tagungen/workshop95/schuur/>.
- [23] M. Randic, *J. Chem. Inf. Comput. Sci.* 35 (1995) 372–382.
- [24] B. Hollas, *J. Math. Chem.* 33 (2) (2003) 91–101, [http://theorie.informatik.uni-ulm.de/Personen/Hollas/papers/ac\\_jmc.pdf](http://theorie.informatik.uni-ulm.de/Personen/Hollas/papers/ac_jmc.pdf).
- [25] G. Moreau, P. Broto, *Nouv. J. Chim.* 4 (1980) 359–360.
- [26] M. Wagener, J. Sadowski, J. Gasteiger, *J. Am. Chem. Soc.* 117 (1995) 7769–7775.
- [27] J. Devillers, D. Domine, *SAR QSAR Environ. Res.* 7 (1997) 195–232.
- [28] M. Randic, *Acta Chim. Slov.* 45 (1998) 239–252.
- [29] E. Estrada, *Internet Electron. J. Mol. Des.* 1 (2002) 360–366, <http://www.biochempress.com>.
- [30] F.R. Burden, *J. Chem. Inf. Comput. Sci.* 29 (1989) 225–227.
- [31] V. Consonni, R. Todeschini, M. Pavan, *J. Chem. Inf. Comput. Sci.* 42 (2002) 682–692.
- [32] V. Consonni, R. Todeschini, M. Pavan, P. Gramatica, *J. Chem. Inf. Comput. Sci.* 42 (2002) 693–705.
- [33] A. Katritzky, L. Mu, *J. Chem. Inf. Comput. Sci.* 36 (1996) 1162–1168.
- [34] A.R. Katritzky, Y. Wang, S. Sild, T. Tamm, M. Karelson, *J. Chem. Inf. Comput. Sci.* 38 (1998) 720–725.
- [35] J. Schuur, *J. Gasteiger, Anal. Chem.* 69 (1997) 2398–2405.
- [36] A.K. Ghose, G.M. Crippen, *J. Comput. Chem.* 7 (1986) 565–577.
- [37] A.K. Ghose, G.M. Crippen, *J. Comput. Chem.* 27 (1987) 21–35.
- [38] I. Moriguchi, S. Hirono, I. Nakagome, H. Hirano, *Chem. Pharm. Bull.* 42 (1994) 976–978.
- [39] I. Moriguchi, S. Hirono, Q. Liu, I. Nakagome, Y. Matsushita, *Chem. Pharm. Bull.* 40 (1992) 127–130.
- [40] Gang Ruan, Si-Shen Feng, Qiu-Tian Li, *J. Controlled Release.* 84 (2002) 151–160.
- [41] J.-L.M. Abboud, R. Notario, *Pure Appl. Chem.* 71 (1999) 645–718.
- [42] K.S. Soppimath, T.M. Aminabhavi, *J. Microencapsulation* 19 (2002) 281–292.
- [43] P.D. Huibers, A.R. Katritzky, *J. Chem. Inf. Comput. Sci.* 38 (1998) 283–292.
- [44] Chongli Zhong, Qinghua Hu, *J. Pharm. Sci.* 92 (2003) 2284–2294.
- [45] F. Deanda, K.M. Smith, Jie Liu, R.S. Pearlman, *Mol. Pharm.* 1 (2004) 23–39.
- [46] G. Selo, *Internet Electron. J. Mol. Des.* 2 (2003) 137–159, <http://www.biochempress.com>.
- [47] L.B. Kier, L.H. Hall, *J. Pharm. Sci.* 72 (1983) 1170–1181.
- [48] A.R. Katritzky, P.D.T. Huibers, V.S. Lobanov, D.O. Shah, M. Karelson, *J. Colloid Interface Sci.* 187 (1997) 113–120.
- [49] A.R. Katritzky, P.D.T. Huibers, D.O. Shah, *J. Colloid Interface Sci.* 193 (1997) 132–136.
- [50] P.T. Huibers, V.S. Lobanov, A.R. Katritzky, D.O. Shah, M. Karelson, *Langmuir* 12 (1996) 1462–1470.
- [51] G. Rücker, C. Rücker, *J. Chem. Inf. Comput. Sci.* 39 (1999) 788–802.
- [52] Hong Yin, Yizeng Liang, Qinnan Hu, *R. J. Data Sci.* 1 (2003) 461–480.
- [53] F.A. de Lima Ribeiro, M.M.C. Ferreira, *J. Mol. Struct. (Theochem)* 663 (2003) 109–126.
- [54] E. Olsen, F. Nielsen, *Molecules* 6 (2001) 370–389.
- [55] H. Wiener, *J. Am. Chem. Soc.* 69 (1947) 17–20.
- [56] L.B. Kier, *Quant. Struct.-Act. Relat.* 4 (1985) 109–116.
- [57] G. Espinosa, D. Yaffe, Y. Cohen, A. Arenas, F. Giralt, *J. Chem. Inf. Comput. Sci.* 40 (2000) 859–879.
- [58] A. Pfennig, A. Schwerin, *Ind. Eng. Chem. Res.* 37 (1998) 3180–3188.
- [59] C.V. Sternling, L.E. Scriven, *AIChE J.* 5 (1959) 514–523.
- [60] M. Henschke, A. Pfennig, *AIChE J.* 45 (1999) 2079–2086.
- [61] G.W. Kauffman, P.C. Jurs, *J. Chem. Inf. Comput. Sci.* 41 (2001) 408–418.
- [62] B. Lucic, I. Basic, D. Nadramija, A. Milicevic, N. Trinajstic, T. Suzuki, R. Petrukhin, M. Karelson, A.R. Katritzky, <http://www.arkat-usa.org/ark/journal/2002/Sunko/DS-381D/381D.htm>.
- [63] M. Randic, Alan R. Liss, In *Molecular Basis of Cancer, Part A: Macromolecular Structure, Carcinogens, and Oncogenes*, New York, 1985, pp. 309–318.