

# **Singlet Oxygen Generation by Porphyrins and Metalloporphyrins Revisited: a Quantitative Structure-Property Relationship (QSPR) Study**

**Andrey A. Buglak<sup>a,\*</sup>, Mikhail A. Filatov<sup>b</sup>, M. Althaf Hussain<sup>c</sup>, Manabu Sugimoto<sup>c</sup>**

<sup>a</sup> Saint-Petersburg State University, 199034 St. Petersburg, Russia

<sup>b</sup> School of Chemical and Pharmaceutical Sciences, Technological University Dublin, City Campus, Kevin Street, Dublin 8, Ireland

<sup>c</sup> Kumamoto University, 860-8555 Kumamoto, Japan

**\* Corresponding Author:**

E-mail: andreybuglak@gmail.com

## **Abstract**

Porphyrins and metalloporphyrins are used as photosensitizers in photocatalysis, photodynamic therapy (PDT), disinfection, degradation of persistent pollutants and other applications. Their mechanism of action involves intersystem crossing to triplet excited state followed by formation of singlet oxygen ( $^1\text{O}_2$ ), which is a highly reactive species and mediates various oxidative processes. The design of advanced sensitizers based on porphyrin compounds have attracted significant attention in recent years. However, it is still difficult to predict the efficiency of singlet oxygen generation for a given structure. Our goal was to develop a quantitative structure-property relationship (QSPR) model for the fast virtual screening and prediction of singlet oxygen quantum yields for porphyrins and metalloporphyrins. We performed QSPR analysis of a dataset containing 32 compounds, including various porphyrins and their analogues (chlorins and bacteriochlorins). Quantum-chemical descriptors were calculated using Density Functional Theory (DFT), namely B3LYP and M062X functionals. Three different machine learning methods were used to develop QSPR models: random forest regression (RFR), support vector regression (SVR), and multiple linear regression (MLR). The optimal QSPR model «structure – singlet oxygen generation quantum yield» obtained using RFR method demonstrated high determination coefficient for the training set ( $R^2 = 0.949$ ) and the highest predicting ability for the test set ( $\text{pred}_R^2 = 0.875$ ). This proves that the developed QSPR method is reliable and

can be directly applied in the studies of singlet oxygen generation both for free base porphyrins and their metal complexes. We believe that QSPR approach developed in this study can be useful for the search of new poprhyrin photosensitizers with enhanced singlet oxygen generation ability.

**Keywords:** porphyrins; photosensitization; singlet oxygen; **Quantitative Structure-Property Relationship**; machine learning.

## 1. Introduction

Organic photosensitizers, efficiently forming long-living triplet excited states and singlet oxygen ( $^1\text{O}_2$ ), found applications in various fields of technology, e.g. photocatalysis, disinfection, photodynamic therapy (PDT) and degradation of persistent pollutants [1]. In recent years there has been a growing number of studies devoted to synthesis and photophysical investigation of novel photosensitizers [2]. However, despite a great number of papers on this topic it is still challenging to design a photosensitizer with predetermined photophysical parameters, in particular absorption in a target region of spectrum, high triplet excited state yield and long lifetime. For this reason, development of structure-activity relationships for predicting photosensitizing ability of organic molecules is crucially important.

Porphyrins represent a unique class of organic photosensitizers due to a combination of easily controllable photophysical properties and synthetic accessibility [3]. Porphyrins have multiple applications including drug delivery [4,5], bioimaging [6,7], and non-linear optics [8,9] to name a few. One of the most intriguing properties of porphyrins is their ability to produce singlet oxygen with high quantum yield under visible light irradiation, which is used for the treatment of skin diseases such as psoriasis [10], port-wine stains [11], as well as different types of cancer like mycosis fungoides (a type of a cutaneous T-cell lymphoma) [12], malignant mesothelioma [13], glioma [14], brain tumor [15], fibrosarcoma [16], colorectal cancer [17], osteosarcoma [18]. Porphyrins and metalloporphyrins are used in antimicrobial PDT [19-21]. Moreover, some of the photosensitizers are used for the photodynamic treatment of such skin diseases as vitiligo, psoriasis, etc.

In cells, porphyrins may participate in photooxidation reactions of electron donors (nucleic acids, peptides, lipids, and other biomolecules). This type of reactions relate to type I photooxidation mechanism:



Radical-anion of the porphyrin ( $\text{P}^{\bullet-}$ ) may also transfer electron to  $\text{O}_2$  with formation of superoxide-anion radical  $\text{O}_2^{\bullet-}$ :

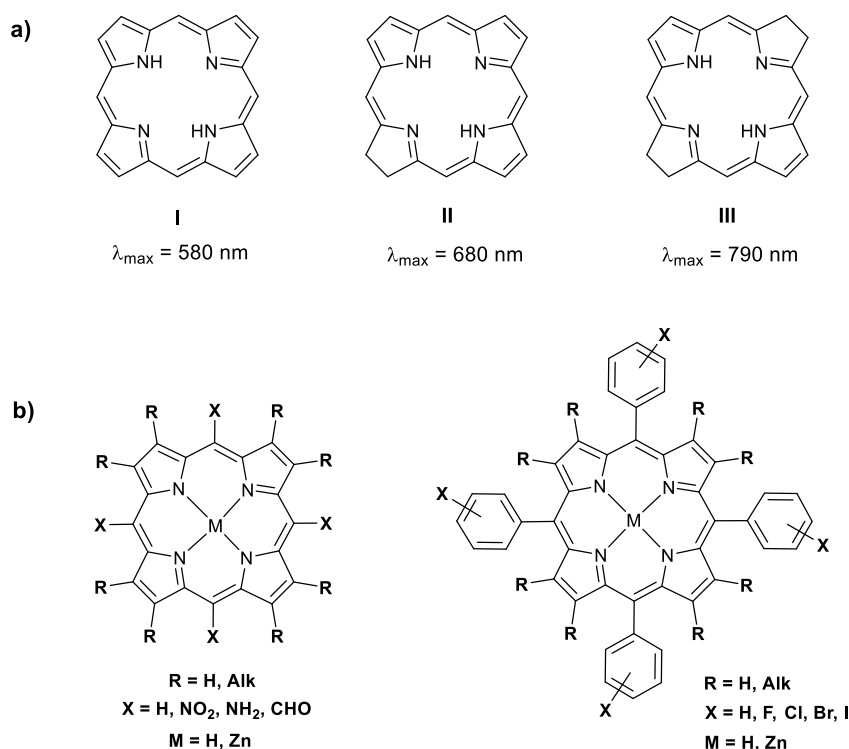


Alternatively, reactive oxygen species (ROS), e.g. singlet oxygen is produced according to type II photooxidation mechanism:



As one can see both types of photooxidation reactions depend on the triplet state formation. We may see that an efficient porphyrin photoagent possesses: 1) good penetration into the cells; 2) high absorption coefficient in the red part of the visible spectrum; 3) high triplet state quantum yield; 4) long triplet state lifetimes; 5) high ROS generation efficiency, in particular, high quantum yield of singlet oxygen generation ( $\Phi_{\Delta}$ ).

The parent compound of porphyrins, porphine (structure I, Fig. 1), exhibits efficient **intersystem crossing (ISC)**. Chemical modification of porphine scaffold allows to optimize its properties for practical applications. For instance, chlorin (II) and bacteriochlorin (III) derivatives possess intense absorption in the near-IR window of tissue and high values of  $\Phi_{\Delta}$ , that makes them ideal photosensitizers [22]. Introduction of peripheral substituents into *meso* and  $\beta$  positions of the macrocycle can be used to modulate the wavelength of absorption maxima within the therapeutic window and tune the solubility. However, it is hard to predict the effect of substituents on ISC and in some cases functionalization (e.g. with strong electron accepting groups) can lead to reduced triplet state yields and lifetimes [23].



**Fig. 1.** a) Porphine (I), chlorin (II) and bacteriochlorin (III) scaffolds. b) General structures of photosensitizers investigated in this work.

Quantitative Structure-Activity Relationship (QSAR) models are commonly used for the prediction of absorption, distribution, metabolism, excretion (ADME), toxicity, and selectivity towards certain cells in pharmacokinetics. To our knowledge QSAR models for the prediction of phototoxicity and for  $^1\text{O}_2$  generation quantum yield prediction are

rare. However, earlier we showed that **Quantitative Structure-Property Relationship (QSPR)** method may be used for the prediction of singlet oxygen generation quantum yield for organic sensitizers [24,25].

The quantum yield of singlet oxygen generation  $\Phi_{\Delta}$  is a ratio between the amount of generated  $^1\text{O}_2$  molecules and the quantity of photons absorbed. Several parameters influence the generation of  $^1\text{O}_2$ : steric and structural effects, triplet state energy ( $E_T$ ), oxidation potential, electronic configuration, solvent polarity, aggregation, and excimer formation [26]. Interaction both with singlet and triplet excited states of a sensitizer can lead to energy transfer. However, for most photosensitizers, the role of ground state molecular oxygen interaction with  $S_1$  state of the photosensitizer is limited to the enhancement of the triplet quantum yield, and the main pathway of  $^1\text{O}_2$  generation is  $\text{O}_2$  interaction with the first excited triplet state,  $T_1$  [26]. Thus,  $\Phi_{\Delta}$  is determined as:

$$\Phi_{\Delta} = \Phi_T p_T^{\text{O}_2} f_T^{\Delta}$$

where  $\Phi_T$  is the triplet state formation quantum yield,  $p_T^{\text{O}_2}$  is the fraction of  $T_1$  state quenched by molecular oxygen,  $f_T^{\Delta}$  is the efficiency of  $^1\text{O}_2$  production during molecular oxygen quenching of  $T_1$ . Because the intramolecular  $T_1$  to  $S_0$  transition is spin-forbidden,  $T_1$  states generally have long lifetimes, allowing almost complete quenching by  $\text{O}_2$  ( $p_T^{\text{O}_2} \approx 1$ ) in air-equilibrated solutions. Incomplete quenching is observed only with triplet states possessing a short lifetime due to high chemical reactivity or fast intersystem crossing (ISC) (i.e.,  $n\pi^*$  triplets), or in the presence of very high concentrations of other quenchers [26].

The higher the rate of physical and chemical quenching of the triplets, the lower the efficiency of energy transfer from the triplet state sensitizer to  $\text{O}_2$  [27]:

$$f_T^{\Delta} = \frac{k_{\text{et}}[\text{O}_2]}{k_{\text{et}}[\text{O}_2] + k_q^T[\text{O}_2] + k_p + k_{\text{ISC}}}$$

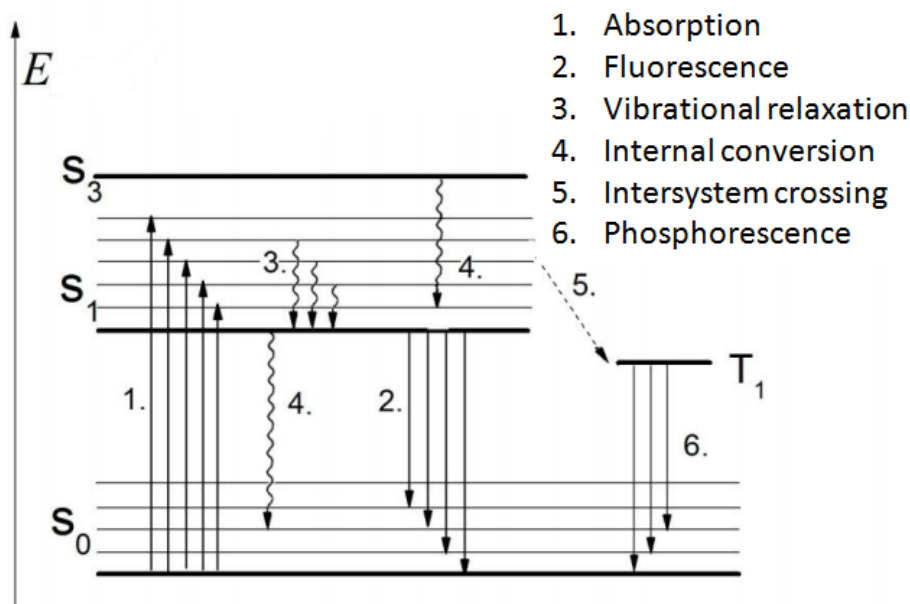
where  $k_{\text{et}}$  is the rate of energy transfer reaction,  $k_q^T$  is the rate constant of other possible processes involving  $^3\text{P}^*$  and  $\text{O}_2$  (e.g. superoxide anion formation, quenching of  $^3\text{P}^*$  by molecular oxygen leading to the ground state of both reactants);  $k_p$  and  $k_{\text{ISC}}$  are the rate constant of radiative decay of the triplet state photosensitizer  $^3\text{P}^*$  and the rate constant of non-radiative decay, respectively. It was found that magnitudes of  $k_q^T$  and  $f_T^{\Delta}$  are inversely correlated, and both parameters exhibit pronounced sensitivity to the oxidation potential [28].

The quantum yield of the triplet state formation depends on the rate of intersystem crossing  $k'_{ISC}$  and the rates of competing processes:

$$\Phi_T = \frac{k'_{ISC}}{k'_{ISC} + k_q + k_{fl} + k_{ic}}$$

where  $k'_{ISC}$  is the rate constant of S-T intersystem crossing,  $k_q$  is the rate constant of physical and chemical quenching of singlet excited state,  $k_{fl}$  is the fluorescence rate constant,  $k_{ic}$  is the internal conversion rate constant (Fig. 2). ISC is governed by spin-orbit coupling (SOC) [29], which is a relativistic effect. SOC is enhanced in the presence of atoms with large nuclei (heavy atoms), such as transition metals and halogens (bromine and iodine). Introduction of heavy atoms into organic chromophores is a common strategy used to access efficient photosensitizers. However, incorporation of heavy atoms can lead to other problems, such as higher costs, toxicity, lower photostability and also a shortened triplet state lifetime [30]. These issues have stimulated much efforts in the field and photosensitizers with efficient ISC in the absence of any heavy atoms are being actively investigated [31,32].

There are qualitative rules, describing ISC. According to El-Sayed rule [33], S-T ISC is efficient between  $^1n\pi^*$  and  $^3\pi\pi^*$ ,  $^1\pi\pi^*$  and  $^3n\pi^*$  states. Another generally known rule is the energy gap law which says that transition probability depends exponentially on the adiabatic energy difference  $\Delta E$ , i.e., the smaller the energy gap the larger the transition probability [34]. In quantum chemistry approximate SOC operators and state-of-the-art theoretical methods are used for evaluation of ISC rates. Since for pharmaceutical purposes it is required to perform virtual screening on large libraries of compounds, calculation of SOC operators for these purposes seems inappropriate. The alternative strategy proposed in this study is the application of the QSPR and machine learning methodology for the prediction of  $\Phi_\Delta$ .



**Fig. 2.** Jablonski diagram, key processes for triplet state formation.

Photosensitizer synthesis and experimental determination of singlet oxygen generation quantum yield is a time- and labor-consuming process. Therefore, preliminary, pre-synthetic screening of sensitizers for their ability to generate  $^1\text{O}_2$  would be of great value. That is why our goal was to build QSPR models predicting the quantum yield of  $^1\text{O}_2$  generation.

To develop a reliable QSPR model, we focused on a series of compounds, combining several classes of functionalized porphyrins and their analogues (Fig. 1b). It includes four sub-groups: 1) tetraphenyl porphyrins derivatives with different substituents in *meso*-phenyl rings and  $\beta$ -positions of pyrrolic rings; 2)  $\beta$ -alkyl substituted porphyrins; 3) derivatives of etioporphyrin and deuteroporphyrin bearing electron donating or accepting groups; 4) chlorin and bacteriochlorin derivatives. In addition, zinc (II) complexes of six structurally different porphyrins were included into the dataset. Previously reported data on  $\Phi_\Delta$  of these compounds in organic solvent were used for analysis. We performed QSPR analysis and optimized the model for predicting photosensitizing ability. We believe that the results of our study will be helpful for the virtual screening of new synthetic porphyrins and metalloporphyrins.

## 2. Methods

All porphyrin compounds studied in this work are poorly soluble in water, but possess good solubility in organic solvents. For this reason, experimental values of  $\Phi_\Delta$  measured in toluene were used for analysis [35]. The activity of compounds, presented as  $\text{Log } \Phi_\Delta$ , was used

as a dependent variable in the QSPR study. The porphyrins were divided into the training (80% of studied compounds) and test sets (20%) using random number generation. The requirements for the maximum and minimum values of  $\text{Log } \Phi_{\Delta}$  in the test set were the following: 1) the maximum  $\text{Log } \Phi_{\Delta}$  value should be less than or equal to the maximum value of the training set; 2) the minimum  $\text{Log } \Phi_{\Delta}$  value should be higher than or equal to the minimum value of the training set.

Two datasets were produced from the original one: 1) to simplify the task 26 free base porphyrins were included in the first dataset; 2) the second one consisted of free base porphyrins and corresponding zinc (II) porphyrins (32 molecules in total). Chemical structures of studied compounds are presented in the Supporting Information (Fig. S1).

## 2.1 Geometry optimization

Conformational analysis of porphyrin molecules was performed using Spartan v. 16 modeling software from Wavefunction, Inc. ([www.wavefun.com](http://www.wavefun.com)). Generation of low-energy conformers was performed using PM6 semi-empirical quantum-chemical method [36]. Geometry optimization was done using Density Functional Theory. We used two different methods for geometry optimization and molecular descriptors calculations. B3LYP [37,38] functional with 6-31G(d,p) basis set for all atoms and LANL2DZ basis set with effective core potential (ECP) [39-41] was used for iodine in Spartan v. 16. We used M06-2X [42,43] with mixed basis set of 6-31G(d,p) for all atoms and SDD basis set with ECP for iodine [44] in Gaussian v. 16 [45]. Optimized porphyrin metalloporphyrin geometries are presented on Fig. S2.

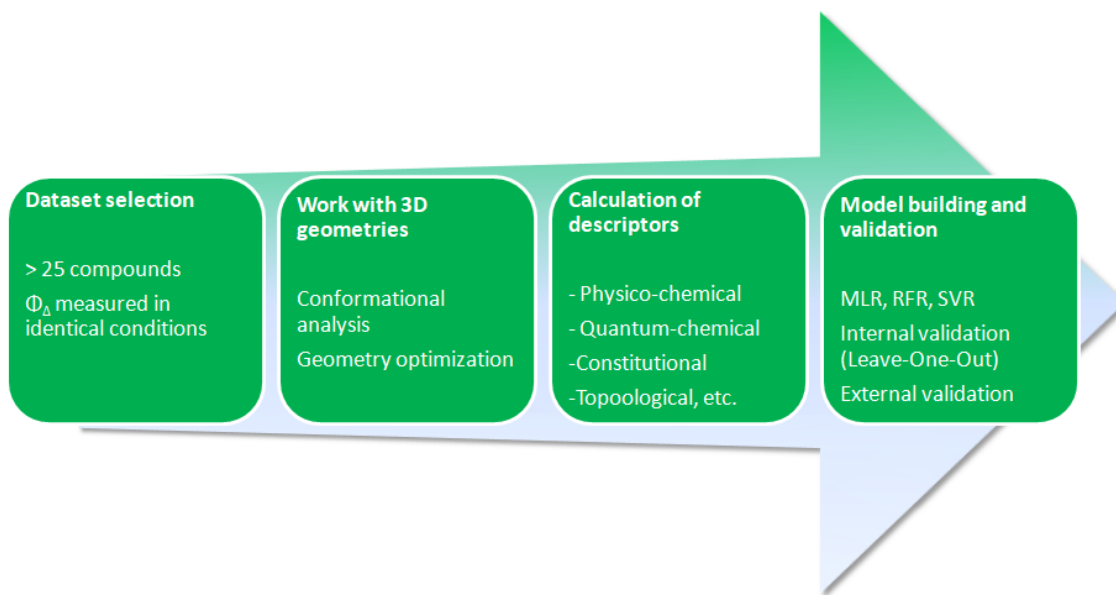
## 2.2 Molecular descriptors

Dragon v. 7.0 descriptors were calculated using online chemical modeling environment OCHEM [46]: a total of 5270 descriptors was calculated. The number of Dragon descriptors was reduced to 156 using Generic Algorithm v. 4.1 (a software developed in Jadavpur University (Kolkata, India) and available at <http://dtclab.webs.com/software-tools>). Twenty six quantum-chemical descriptors like dipole moment, frontier molecular orbital energies, HOMO-LUMO gap, electronegativity, polarizability, reorganization energies were obtained at B3LYP/6-31G(d,p),LANL2DZ and M06-2X/6-31G(d,p),SDD level of theory for the gas phase. Several parameters like singlet and triplet vertical excitation energies, and oscillator strength of  $S_0 \rightarrow S_1$  and  $S_0 \rightarrow S_2$  electron transitions, as well as S-T energy gap, were calculated with Time-Dependent Density Functional Theory (TDDFT) and M06-2X functional.



### 2.3 Machine learning techniques. Model search

Machine learning (ML) was performed using scikit-learn library of Python programming language. We used three different ML methods: support vector regression (SVR), multiple linear regression (MLR), and random forest regression (RFR) (Fig. 3).



**Fig. 3.** Modeling workflow.

The required MLR equation should have the following form:

$$y = c + a_1 * x_1 + \dots + a_n * x_n$$

where  $y$  is the dependent variable  $\text{Log } \Phi_{\Delta}$ ,  $c$  is a regression constant,  $a_1$  and  $a_n$  are regression coefficients,  $x_1$  and  $x_n$  are independent variables. The equation obtained should contain a maximum of four descriptors for a training set of 21 compounds and five descriptors for a training set of 26 compounds (at least five compounds per molecular descriptor) to avoid over-fitting.

MLR model search was performed using Genetic Algorithm v. 4.1. SVR and RFR models search was done in scikit-learn using grid-search method and 5-fold cross validation. SVR model search was done by varying three parameters:  $C$ , epsilon, and kernel (linear, polynomial, sigmoid or radial basis function). RFR search was performed by changing the values of two parameters: number of estimators (trees) and maximal depth. Other parameters were used by default.

We performed standard-scaling of descriptor values for SVR models. The idea of standard-scaling is to make the values of each descriptor in the dataset have zero mean and unit-variance, according to the expression:

$$\frac{X - \text{mean}(X)}{\text{standard\_deviation}(X)}$$

## 2.4 Descriptors contribution to the model

We estimated the contribution of each descriptor to the MLR models using following equation:

$$\alpha(x_1) = \frac{R^2(x_1)}{R^2(x_1) + \dots + R^2(x_n)} \times 100\%$$

where  $\alpha(x_1)$  is the relative contribution of the descriptor  $x_1$  to the model with several descriptors,  $R^2(x_n)$  is the determination coefficient of the  $n$ th descriptor towards  $\text{Log } \Phi_{\Delta}$ .

Relative contribution of descriptors in SVR and RFR models was evaluated by using the facilities implemented in these methods: feature importance was evaluated in RFR; in SVR with linear kernel coefficients of the descriptors were used for relative contribution calculation according to the following formula:

$$\alpha(x_1) = \frac{\text{coef}(x_1)}{\sum \text{coef}(" + ") - \sum \text{coef}(" - ")} \times 100\%$$

where  $\text{coef}(x_1)$  is the value of the coefficient for descriptor  $x_1$ ;  $\sum \text{coef}(" + ")$  and  $\sum \text{coef}(" - ")$  are the sums of coefficients for all positive and negative coefficients, respectively.

## 2.5 Statistical parameters

Best QSPR models were selected on the basis of statistical parameters such as  $R^2$  (coefficient of determination of the training set),  $q^2$  (internally cross-validated  $R^2$  using the leave-one-out (LOO) method), and  $\text{pred}_R^2$  (predictive  $R^2$  for the test set of compounds).  $Q^2$  and  $\text{pred}_R^2$  were used for model validation and comparison. QSPR models obtained were validated and tested for predicting ability using a test set of compounds.

To calculate  $q^2$  parameter each molecule in the training set was excluded once and  $\text{Log } \Phi_{\Delta}$  of the excluded molecule was predicted by using the model developed by the remaining compounds.  $Q^2$  describes the internal stability of a model and was calculated using the following formula:

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2}$$

where  $y_i$  and  $\hat{y}_i$  are the actual and predicted  $\text{Log } \Phi_{\Delta}$  value of the  $i$ th molecule in the training set, respectively;  $y_{mean}$  is the average  $\text{Log } \Phi_{\Delta}$  of all compounds in the training set.

For external validation, the activity of each compound in the test set was predicted using the model developed by the training set.  $\text{Pred}_R^2$  indicates the predictive ability of the model.  $\text{Pred}_R^2$  was calculated according to the equation:

$$pred\_R^2 = 1 - \frac{\sum(y_{act} - y_{pred})^2}{\sum(y_{act} - y_{mean})^2}$$

where  $y_{act}$  and  $y_{pred}$  are the actual and predicted activity of the  $i$ th compound in the test set, respectively;  $y_{mean}$  is the average Log  $\Phi_{\Delta}$  of all molecules in the training set. Both summations are over all compounds in the test set.

### 3. Results and Discussion

To simplify the task we excluded metalloporphyrins and started with the analysis of a dataset consisting of 26 free base porphyrins [35]. Most of the compounds in this dataset are derivatives of tetraphenylporphyrin (TPP), containing different substituents in *meso*-phenyl groups. We also included various representatives of  $\beta$ -alkyl-substituted porphyrins as well as derivatives of chlorin and bacteriochlorin. We used three different machine learning methods in our analysis: random forest (RFR), support vector regression (SVR), and multiple linear regression (MLR). Statistical parameters of the models obtained are summarized in Table 1.

**Table 1.** Statistical parameters of three models predicting the quantum yield of singlet oxygen generation by porphyrins: support vector regression (SVR), multiple linear regression (MLR), and random forest (RFR).

Parameter	SVR	MLR	RFR
R	0.984	0.929	0.981
R <sup>2</sup>	0.968	0.863	0.962
q <sup>2</sup>	0.561	0.733	0.787
pred_R <sup>2</sup>	0.694	0.876	0.869
RMSE	0.037	0.043	0.029
max $\Phi_{\Delta}$ error ( $\Phi_{\Delta}$ in %)	12.3	15.8	13.5
RMS $\Phi_{\Delta}$ error ( $\Phi_{\Delta}$ in %)	3.9	6.5	4.5

A QSPR model is considered predictive if the following conditions are satisfied:  $R^2 > 0.6$ ,  $q^2 > 0.6$ , and  $pred\_R^2 > 0.5$  [47]. As can be seen from Table 1, all three models are sufficient. The SVR model has the highest values of R and R<sup>2</sup> as well as the lowest values of maximum  $\Phi_{\Delta}$  error (12.3 %) and root mean square  $\Phi_{\Delta}$  error (3.9%). The MLR model has the highest value of the most important statistical parameter  $pred\_R^2$ . The RFR model is perhaps the most internally stable one ( $q^2 = 0.787$ ) and possesses the lowest value of root mean square error (0.029).

#### 3.1 Model 1 (support vector regression)

Let’s regard the models obtained. Thus, we built a SVR model with linear kernel (Model 1). The model possesses constants C and epsilon both equal to 0.0025. Linear kernel makes it

possible to evaluate descriptors contribution. The SVR model uses a full set of 182 descriptors, but the two most influential descriptors are *MATS1m* (relative contribution -1.7%) and *HATS4m* (1.5%). *MATS1m* is a Moran autocorrelation lag 1 / weighted by atomic masses (2D autocorrelation descriptor) [48]. The descriptor is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . N-Methyloctaethylporphyrin has the highest value of *MATS1m* equal to 0.063, while tetra(4-bromophenyl)porphyrin ((p-Br)<sub>4</sub>-TPP) has the lowest value equal to -0.002. In general, halogenated molecules have low values of the descriptor while halogen-free molecules have high values of *MATS1m*. That means that halogenation is favorable for high  $\Phi_{\Delta}$ , which is a well-known effect [49]. *HATS4m* is a leverage-weighted autocorrelation of lag 4 / weighted by atomic masses (GETAWAY descriptor) [50,51]. The descriptor is directly proportional to  $\text{Log } \Phi_{\Delta}$ . The value of *HATS3m* is equal to zero for 10 compounds presented in the dataset (Table 2), while tetra(4-iodophenyl)porphyrin ((p-I)<sub>4</sub>-TPP) has the highest value of the descriptor equal to 0.134. Like *MATS1m*, *HATS3m* values significantly vary from halogenated compounds to halogen-free compounds.

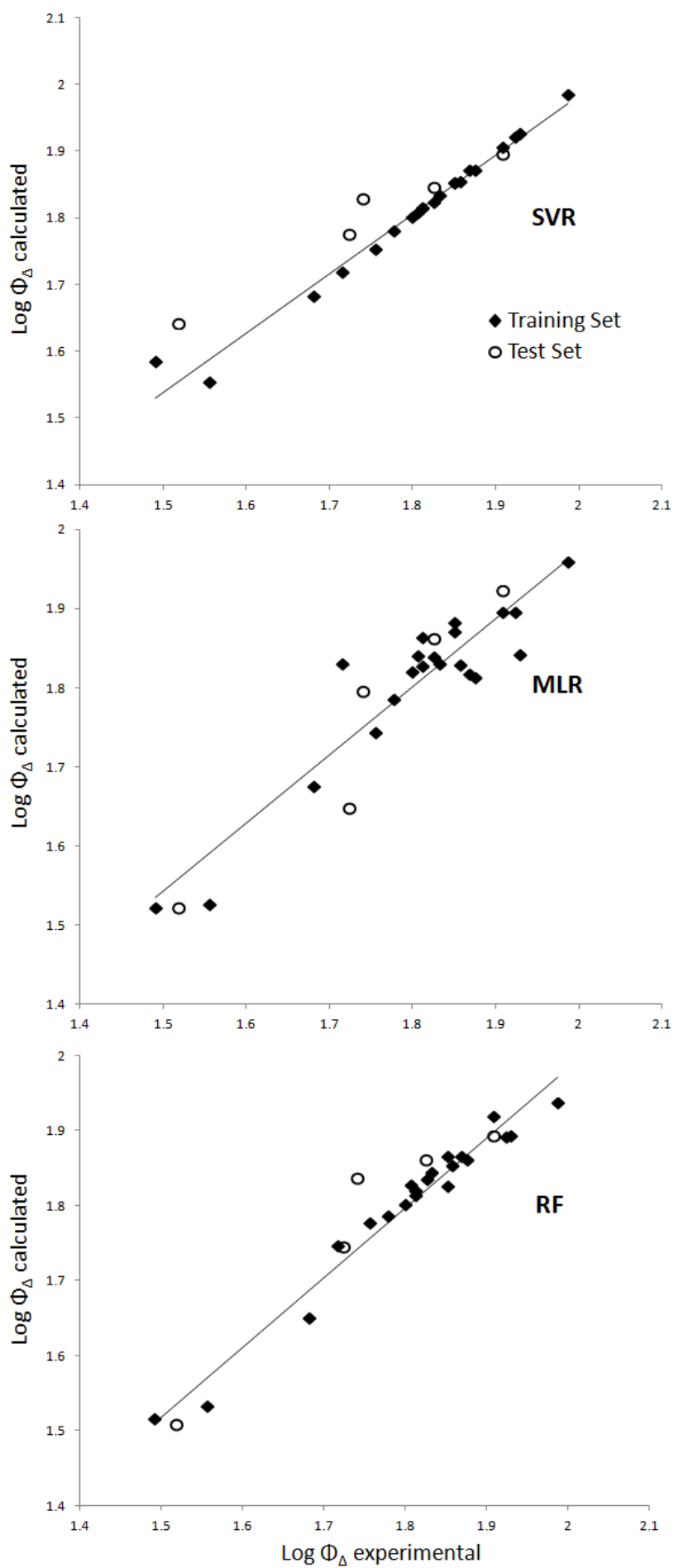
**Table 2.** Values of two main descriptors used in Model 1 (support vector regression).

Compound	MATS1m	HATS4m
(m-Cl) <sub>4</sub> -TPP	0.005	0.109
(o-Cl) <sub>4</sub> -TPP	0.005	0.112
(o-F) <sub>4</sub> -TPP	0.027	0.072
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	-0.002	0.095
(p-Cl) <sub>4</sub> -TPP	0.005	0.062
(p-F) <sub>4</sub> -TPP	0.027	0.049
(p-I) <sub>4</sub> -TPP	0	0.134
2,3,7,8-Tetraethylporphyrin	0.06	0.048
2,3,12,13-Tetraethylporphyrin	0.06	0
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	0.06	0.045
5,10,15-NO <sub>2</sub> -etioporphyrin I	0.061	0.065
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	0.06	0.057
5,15-NO <sub>2</sub> -etioporphyrin I	0.06	0
5-CHO-Octaethylporphyrin	0.051	0
5-NH <sub>2</sub> -etioporphyrin I	0.048	0
5-NO <sub>2</sub> -etioporphyrin I	0.059	0
Deuteroporphyrin-IX-DME <sup>test</sup>	0.052	0
Etioporphyrin I	0.059	0
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	0.063	0
Octaethylchlorin	0.058	0.032
Octaethylporphyrin	0.057	0
Octamethyl-TPP	0.06	0.027
Porphine <sup>test</sup>	0.048	0
Tetraphenylchlorin	0.059	0.036
Tetraphenyltetrahydroporphyrin	0.062	0.037
TPP	0.056	0

<sup>test</sup> designates the compounds belonging to the test set

Experimental values of  $\text{Log } \Phi_{\Delta}$  and values predicted by Model 1 for each porphyrin compound are presented in Table S1 and in Fig. 4.

We randomly divided the dataset into the training and testing sets four times and tested the models to additionally confirm good quality of the model obtained. Thus, key parameters had the following value ranges after additional dataset splitting:  $R^2 = 0.952-0.993$ ,  $\text{pred}_R^2 = 0.610-0.691$ ,  $q^2 = 0.524-0.734$ , and  $\text{RMSE} = 0.019-0.037$ . These results show that Model 1 is tolerant to alternative dataset splitting, possesses high predicting ability and internal stability.



**Fig. 4.** Experimental vs. predicted  $\text{Log } \Phi_{\Delta}$  values for 26 free base porphyrins.

### 3.2 Model 2(multiple linear regression)

Next, we obtained a linear regression equation predicting the  $\Phi_{\Delta}$  of 26 porphyrins. The equation is presented below:

$$\text{Log } \Phi_{\Delta} = 1.829(+/-0.016) - 0.400(+/-0.228) \text{ MATS1e} - 0.077(+/-0.026) * \text{Eig02\_EA(dm)} + 0.345(+/-0.146) * \text{HATS3m} - 0.271(+/-0.112) * \text{S1\_fosc} \quad (\text{Model 2})$$

*MATS1e* (relative contribution -43.4%) is a Moran autocorrelation lag 1 / weighted by atomic Sanderson electronegativities [48]. *MATS1e* is inversely correlated with Log  $\Phi_{\Delta}$ . Among the studied compounds 5,10,15-NO<sub>2</sub>-etioporphyrin I has the highest value of *MATS1e* equal to 0.279, while porphine has the value of the descriptor equal to -0.086 (Table 3). Electronegative nitrogen atoms are evenly distributed in porphine, which makes its *MATS1e* value low, while in 5,10,15-NO<sub>2</sub>-etioporphyrin I, 5,10-NO<sub>2</sub>-etioporphyrin I, and 5,15-NO<sub>2</sub>-etioporphyrin I introduction of nitro groups results in asymmetrical distribution. Apparently, the main point is that *MATS1e* descriptor allows to distinguish molecules bearing electron accepting groups, e.g. nitro groups, from molecules lacking such groups.

**Table 3.** Molecular descriptor values used in Model 2.

compound	MATS1e	Eig02_EA(dm)	HATS3m	S1_fosc
(m-Cl) <sub>4</sub> -TPP	-0.04	0	0.155	0.0103
(o-Cl) <sub>4</sub> -TPP	-0.04	0	0.146	0.0015
(o-F) <sub>4</sub> -TPP	-0.026	0	0.095	0.0047
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	-0.048	0	0.231	0.0203
(p-Cl) <sub>4</sub> -TPP	-0.04	0	0.125	0.0189
(p-F) <sub>4</sub> -TPP	-0.026	0	0.086	0.018
(p-I) <sub>4</sub> -TPP	-0.037	0	0.349	0.0209
2,3,7,8-Tetraethylporphyrin	-0.001	0	0.038	0.0062
2,3,12,13-Tetraethylporphyrin	-0.001	0	0	0.002
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	-0.001	0	0.039	0.0115
5,10,15-NO <sub>2</sub> -etioporphyrin I	0.279	2.61	0.04	0.0121
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	0.253	2.61	0.038	0.0117
5,15-NO <sub>2</sub> -etioporphyrin I	0.253	2.61	0	0.0168
5-CHO-Octaethylporphyrin	0.004	0	0	0.0282
5-NH <sub>2</sub> -etioporphyrin I	-0.045	0	0	0.0165
5-NO <sub>2</sub> -etioporphyrin I	0.202	0	0	0.0155
Deuteroporphyrin-IX-DME <sup>test</sup>	-0.026	2.44	0	0.0109
Etioporphyrin I	0.017	0	0	0.016
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	0.073	0	0	0.0131
Octaethylchlorin	0.033	0	0.024	0.1429
Octaethylporphyrin	0.029	0	0	0.0159
Octamethyl-TPP	0.016	0	0.035	0.0131
Porphine <sup>test</sup>	-0.086	0	0	0.0022
Tetraphenylchlorin	-0.01	0	0.047	0.0771
Tetraphenyltetrahydroporphyrin	-0.004	0.6	0.048	0.4625
TPP	-0.016	0	0	0.0165

<sup>test</sup> designates the compounds belonging to the test set

*Eig02\_EA(dm)* (relative contribution -37.2%) is an eigenvalue No 2 from the edge adjacency matrix weighted by dipole moment. The descriptor is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . 5,10,15-NO<sub>2</sub>-etioporphyrin I, 5,10-NO<sub>2</sub>-etioporphyrin I, and 5,15-NO<sub>2</sub>-etioporphyrin I have the highest values of *Eig02\_EA(dm)* equal to 2.61, while for most of other compounds it is equal to 0. Similarly to *MATS1e*, *Eig02\_EA(dm)* allows to distinguish molecules bearing electron accepting groups. High values of *Eig02\_EA(dm)* for nitro-substituted porphyrins allow to predict that such compounds have lower values of <sup>1</sup>O<sub>2</sub> generation quantum yield than other porphyrins in the dataset.

*HATS3m* (relative contribution 16.1%) is a leverage-weighted autocorrelation of lag 3 / weighted by atomic masses (GETAWAY descriptor) [50,51]. High *HATS3m* is favorable for high  $\text{Log } \Phi_{\Delta}$ . The value of *HATS3m* is equal to zero for 11 compounds presented in the dataset, while (p-I)<sub>4</sub>-TPP has the highest value of the descriptor equal to 0.349. Apparently, introduction of halogens, especially bromine and iodine, increases  $\Phi_{\Delta}$  values, which is a well-known effect [52,53].

*SI\_fosc* (relative contribution -3.2%) is an oscillator strength for the first electronic transition S<sub>0</sub>→S<sub>1</sub> calculated with M06-2X/6-31G(d,p),SDD method in the gas phase. The descriptor is inversely correlated with  $\text{Log } \Phi_{\Delta}$  and has a minor contribution to Model 2. Tetraphenyltetrahydroporphyrin has the highest value of *SI\_fosc* equal to 0.4625 while (o-Cl)<sub>4</sub>-TPP has the lowest value of the descriptor (0.015). According to Kasha's rule, emission always occurs from the lowest excited state of the same multiplicity (S<sub>1</sub>). The oscillator strength of a transition between the ground state and the excited state is related directly to the intensity of emission [54,55]. Apparently, molecules with higher *SI\_fosc* possess higher emission rate and lower intersystem crossing rate (a competing process), which results in the decrease of both  $\Phi_{\text{T}}$  and  $\Phi_{\Delta}$ .

Experimental and predicted by the MLR model  $\Phi_{\Delta}$  values for each porphyrin compound are presented in Table S2 and in Fig. 4.

We randomly splitted the dataset four times and tested the model performance in each case. Key statistical parameters had the following value ranges:  $R^2 = 0.844-0.922$ ,  $\text{pred}_R^2 = 0.646-0.905$ ,  $q^2 = 0.495-0.827$ , and  $\text{RMSE} = 0.034-0.040$ . As one can see, statistical parameters after alternative dataset splitting are sufficient.

### 3.3 Model 3 (random forest regression)

Next, we have developed a random forest regression model (Model 3). The model selected 54 descriptors out of the total of 182 descriptors; eight trees with a maximal depth equal to 10 were used by the RFR model. Two descriptors have a relative contribution more than 10%. We regarded these descriptors in details.



*nDB* (relative contribution -37.9%) is a number of double bonds in the molecule (constitutional descriptor). *nDB* is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . 5,10,15-NO<sub>2</sub>-etioporphyrin I and tetraphenyltetrahydroporphyrin have the highest value of this descriptor equal to 8 while octaethylchlorin and tetraphenylchlorin have the lowest value of *nDB* equal to 1 (Table 4). Two possible interpretations of the role of *nDB* descriptor can be proposed. First, increasing the number of the double bonds leads to extension of the  $\pi$  system, and consequently to higher values of molecular extinction coefficients and red shift of the absorption maxima. Secondly, singlet oxygen may take part in [4+2] cycloaddition reaction with porphyrin molecules. Such photooxidation processes involving singlet oxygen are common for electron-rich aromatic compounds [26], including  $\pi$ -extended porphyrins [56,57]. Thus, molecules having higher number of double bonds are more likely to react with <sup>1</sup>O<sub>2</sub> and are expected to have lower values of  $\Phi_{\Delta}$ . Porphyrins chemically quench <sup>1</sup>O<sub>2</sub> with the rate constants lying in the range  $1 \times 10^4 - 1 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$  [58].

**Table 4.** Values of two main descriptors used in Model 3 (random forest).

compound	nDB	SpDiam_EA(bo)
(m-Cl) <sub>4</sub> -TPP	2	7.57
(o-Cl) <sub>4</sub> -TPP	2	7.58
(o-F) <sub>4</sub> -TPP	2	7.58
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	2	7.56
(p-Cl) <sub>4</sub> -TPP	2	7.56
(p-F) <sub>4</sub> -TPP	2	7.56
(p-I) <sub>4</sub> -TPP	2	7.56
2,3,7,8-Tetraethylporphyrin	2	7.54
2,3,12,13-Tetraethylporphyrin	2	7.54
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	2	7.53
5,10,15-NO <sub>2</sub> -etioporphyrin I	8	7.78
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	6	7.73
5,15-NO <sub>2</sub> -etioporphyrin I	6	7.71
5-CHO-Octaethylporphyrin	3	7.65
5-NH <sub>2</sub> -etioporphyrin I	2	7.63
5-NO <sub>2</sub> -etioporphyrin I	4	7.69
Deuteroporphyrin-IX-DME <sup>test</sup>	4	7.53
Etioporphyrin I	2	7.56
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	2	7.7
Octaethylchlorin	1	7.56
Octaethylporphyrin	2	7.57
Octamethyl-TPP	2	7.77
Porphine <sup>test</sup>	2	7.22
Tetraphenylchlorin	1	7.55
Tetraphenyltetrahydroporphyrin	8	7.79
TPP	2	7.56

<sup>test</sup> designates the compounds belonging to the test set

*SpDiam\_EA(bo)* (relative contribution -17.8%) is a spectral diameter from edge adjacency matrix weighted by bond order [59-61]. The descriptor is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . Tetraphenyltetrahydroporphyrin has the highest value of the descriptor equal to 7.79,

while porphine has the lowest value equal to 7.22 (Table 4). Bulky substituents are known to hinder the interaction between triplet excited states of porphyrin molecules and ground state molecular oxygen, resulting in lower  $^1\text{O}_2$  quantum yields [62].

Experimental and predicted by random forest model (Model 3)  $\text{Log } \Phi_{\Delta}$  values for each porphyrin compound are presented in Table S3 and in Fig. 4.

After getting a proof-of-principle for QSPR application in the study of singlet oxygen generation by free base porphyrins, we decided to extend this method to a larger dataset of molecules, which includes corresponding metalloporphyrins. A dataset consisting of twenty-six free base porphyrins and six zinc (II) porphyrins was analyzed. Statistical parameters of the models obtained are summarized in Table 5.

**Table 5.** Statistical parameters of support vector regression (SVR), multiple linear regression (MLR), and random forest (RFR) models predicting the quantum yield of singlet oxygen generation by porphyrins and metalloporphyrins.

Parameter	SVR	MLR	RFR
R	0.995	0.943	0.974
$R^2$	0.991	0.889	0.949
$q^2$	0.842	0.744	0.619
pred_ $R^2$	0.801	0.866	0.875
RMSE	0.029	0.045	0.034
max $\Phi_{\Delta}$ error ( $\Phi_{\Delta}$ in %)	14.5	21.9	9.2
RMS $\Phi_{\Delta}$ error ( $\Phi_{\Delta}$ in %)	3.5	7.0	4.1

As one can see from Table 5, the statistical parameters of all three models are sufficient. The SVR model has the best values of R,  $R^2$ ,  $q^2$ , root mean square error, and root mean square  $\Phi_{\Delta}$  error. The RFR model has the best values of maximum  $\Phi_{\Delta}$  error and pred\_ $R^2$ . Pred\_ $R^2$  is probably the most important parameter evaluating the predictive ability of the model towards the test set. We suppose that RFR model is the optimal QSPR model «structure – singlet oxygen generation quantum yield» since it possesses high determination coefficient for the training set ( $R^2 = 0.949$ ) and the highest predicting ability towards the test set (pred\_ $R^2 = 0.875$ ).

Key statistical parameters after alternative splitting had the following value ranges, according to Model 3:  $R^2 = 0.926-0.961$ , pred\_ $R^2 = 0.611-0.812$ ,  $q^2 = 0.620-0.783$ , and RMSE = 0.023-0.035. As one can see, the statistical parameters after alternative dataset splitting are sufficient. However, SVR model (Model 1) possesses highest determination coefficients; MLR model possesses the highest predicting ability towards the test sets, but in one of the splittings  $q^2$  is barely sufficient (0.495). Nevertheless, we suppose that all three models possess satisfactory tolerance to alternative dataset splitting.

### 3.4 Model 4 (SVR)

We built a support vector regression (SVR) model with linear kernel (Model 4). The model possesses constants C and epsilon both equal to 0.005. Linear kernel allows to evaluate relative contributions of the descriptors in the model. The SVR model uses a full set of 182 descriptors, but the two most influential are *SI\_energy* (relative contribution 3.1%) and *Eig02\_EA(dm)* (-2.5%). *SI\_energy* is an energy of the first electronic transition  $S_0 \rightarrow S_1$  calculated with M06-2X/6-31G(d,p) method for the gas phase. The descriptor is directly proportional to  $\text{Log } \Phi_{\Delta}$ . Zn-Etioporphyrin possesses the highest value of  $S_1$  state energy (2.437 eV) while octamethyl-TPP has the lowest value of *SI\_energy* equal to 2.036 eV. Hence, the importance of  $S_0 \rightarrow S_1$  transition arises again: see Model 2 for comparison. We suppose that the role of  $S_1$  state energy reveals a well-known phenomenon that higher  $S_1$  energy is proportional to a lower rate of chemical quenching of  $^1\text{O}_2$  [26]. the compounds capable of  $^1\text{O}_2$  quenching either do not generate singlet oxygen or do it less efficiently. *Eig02\_EA(dm)* is an eigenvalue No 2 from edge adjacency matrix weighted by dipole moment, which was previously used in Model 2 and allows to distinguish molecules with strong electron accepting groups.

**Table 6.** Values of the two main descriptors used in Model 4 (support vector regression).

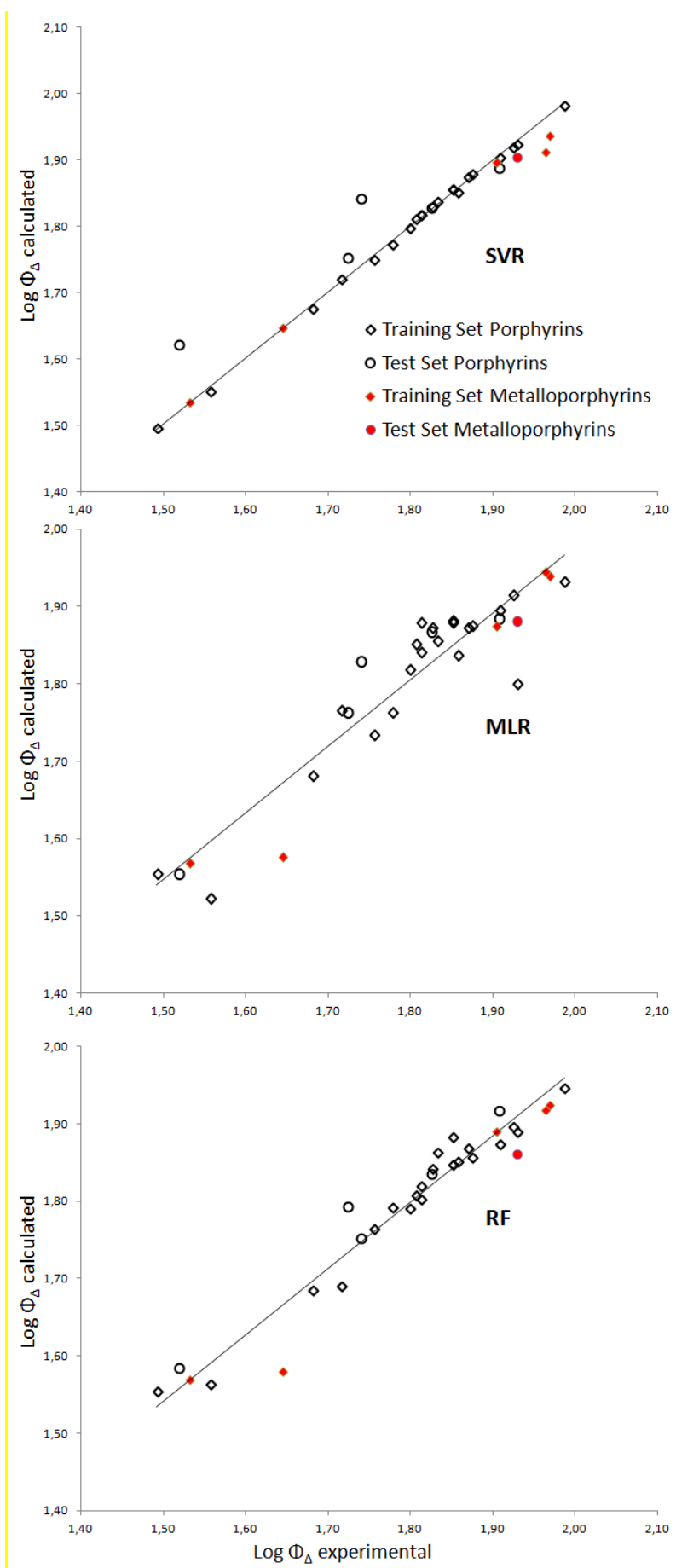
Compound	SI_energy, eV	Eig02_EA(dm)
(m-Cl) <sub>4</sub> -TPP	2.179	0
(o-Cl) <sub>4</sub> -TPP	2.209	0
(o-F) <sub>4</sub> -TPP	2.180	0
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	2.161	0
(p-Cl) <sub>4</sub> -TPP	2.162	0
(p-F) <sub>4</sub> -TPP	2.161	0
(p-I) <sub>4</sub> -TPP	2.162	0
2,3,7,8-Tetraethylporphyrin	2.298	0
2,3,12,13-Tetraethylporphyrin	2.279	0
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	2.267	0
5,10,15-NO <sub>2</sub> -etioporphyrin I	2.240	2.61
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	2.248	2.61
5,15-NO <sub>2</sub> -etioporphyrin I	2.250	2.61
5-CHO-Octaethylporphyrin	2.196	0
5-NH <sub>2</sub> -etioporphyrin I	2.303	0
5-NO <sub>2</sub> -etioporphyrin I	2.254	0
Deuteroporphyrin-IX-DME <sup>test</sup>	2.263	2.44
Etioporphyrin I	2.253	0
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	2.196	0
Octaethylchlorin	2.189	0
Octaethylporphyrin	2.253	0
Octamethyl-TPP	2.036	0
Porphine <sup>test</sup>	2.294	0
Tetraphenylchlorin	2.197	0
Tetraphenyltetrahydroporphyrin	2.313	0.6
TPP	2.165	0
Zn-(5,10-NO <sub>2</sub> )etioporphyrin I	2.429	2.61
Zn-(5,15-NO <sub>2</sub> )etioporphyrin I	2.430	2.61
Zn-(o-F) <sub>4</sub> -TPP	2.375	0

Zn-Etioporphyrin	2.437	0
Zn-Octaethylporphyrin <sup>test</sup>	2.433	0
Zn-TPP	2.355	0

---

<sup>test</sup> designates the compounds belonging to the test set

Experimental values of  $\text{Log } \Phi_{\Delta}$  and values predicted by Model 4 for each porphyrin and metalloporphyrin compound are presented in Table S4 and in Fig. 5.



**Fig. 5.** Experimental vs. predicted  $\text{Log } \Phi_{\Delta}$  for the dataset combining free base porphyrins and metalloporphyrins.

We randomly divided the dataset into the training and testing sets four times and tested the models to additionally confirm good quality of the Model 4. Main statistical parameters had the following value ranges after additional dataset splitting:  $R^2 = 0.976-0.985$ ,  $\text{pred}_R^2 = 0.567-0.881$ ,  $q^2 = 0.654-0.799$ , and  $\text{RMSE} = 0.016-0.031$ . These results show that Model 4 is tolerant to alternative splitting and possesses high predicting ability and stability: determination coefficients  $R^2$  are especially high.

### 3.5 Model 5 (MLR)

We obtained a linear regression equation predicting  $\text{Log } \Phi_{\Delta}$  of 32 compounds. The equation is following:

$$\begin{aligned} \text{Log } \Phi_{\Delta} = & -0.470(+/-0.682) - 0.052(+/-0.024 * \text{Eig02\_EA}(dm) - 0.660(+/-0.229) * \text{MATSIe} \\ & + 0.362(+/-0.138) * \text{SpMax4\_Bh}(p) + 0.112(+/-0.175) * \text{S1\_fosc} + \\ & 0.343(+/-0.096) * \text{HL\_Gap} \quad (\text{Model 5}) \end{aligned}$$

*Eig02\_EA(dm)* (relative contribution -37.2%) is an eigenvalue No 2 from edge adjacency matrix weighted by dipole moment. The descriptor was used in Model 2. It is inversely correlated with  $\text{Log } \Phi$ . *Eig02\_EA(dm)* allows to distinguish molecules with electron accepting groups, which have a negative influence on  $^1\text{O}_2$  generation quantum yields.

*MATSIe* (relative contribution -36.9%) is a Moran autocorrelation lag 1 / weighted by atomic Sanderson electronegativities [48]. *MATSIe* is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . As well as *Eig02\_EA(dm)*, *MATSIe* allows to distinguish molecules bearing electron accepting groups.

**Table 7.** Molecular descriptor values used in Model 5 (MLR).

compound	MATSIe	SpMax4_Bh(p)	S1_fosc	HL_Gap, eV
(m-Cl) <sub>4</sub> -TPP	-0.04	3.86	0.0103	2.74
(o-Cl) <sub>4</sub> -TPP	-0.04	3.86	0.0015	2.8
(o-F) <sub>4</sub> -TPP	-0.026	3.84	0.0047	2.75
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	-0.048	3.87	0.0203	2.68
(p-Cl) <sub>4</sub> -TPP	-0.04	3.86	0.0189	2.69
(p-F) <sub>4</sub> -TPP	-0.026	3.84	0.018	2.74
(p-I) <sub>4</sub> -TPP	-0.037	4.02	0.0209	2.68
2,3,7,8-Tetraethylporphyrin	-0.001	3.63	0.0062	2.93
2,3,12,13-Tetraethylporphyrin	-0.001	3.62	0.002	2.9
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	-0.001	3.67	0.0115	2.95
5,10,15-NO <sub>2</sub> -etioporphyrin I	0.279	3.71	0.0121	2.82
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	0.253	3.71	0.0117	2.92
5,15-NO <sub>2</sub> -etioporphyrin I	0.253	3.71	0.0168	2.86
5-CHO-Octaethylporphyrin	0.004	3.74	0.0282	2.72
5-NH <sub>2</sub> -etioporphyrin I	-0.045	3.71	0.0165	2.61
5-NO <sub>2</sub> -etioporphyrin I	0.202	3.71	0.0155	2.89
Deuteroporphyrin-IX-DME <sup>test</sup>	-0.026	3.68	0.0109	2.94
Etioporphyrin I	0.017	3.71	0.016	2.94
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	0.073	3.74	0.0131	2.89
Octaethylchlorin	0.033	3.73	0.1429	2.59

Octaethylporphyrin	0.029	3.74	0.0159	2.94
Octamethyl-TPP	0.016	3.84	0.0131	2.49
Porphine <sup>test</sup>	-0.086	3.54	0.0022	2.91
Tetraphenylchlorin	-0.01	3.84	0.0771	2.64
Tetraphenyltetrahydroporphyrin	-0.004	3.84	0.4625	2.15
TPP	-0.016	3.84	0.0165	2.69
Zn-(5,10-NO <sub>2</sub> )etioporphyrin I	0.253	3.71	0.0271	2.92
Zn-(5,15-NO <sub>2</sub> )etioporphyrin I	0.253	3.71	0.0528	2.89
Zn-(o-F) <sub>4</sub> -TPP	-0.026	3.84	0.0026	2.92
Zn-Etioporphyrin	0.017	3.71	0.0355	2.94
Zn-Octaethylporphyrin <sup>test</sup>	0.029	3.74	0.0349	2.95
Zn-TPP	-0.016	3.84	0.0146	2.95

<sup>test</sup> designates the compounds belonging to the test set

*SpMax4\_Bh(p)* (relative contribution 16.3%) is a largest eigenvalue No 4 of Burden matrix weighted by polarizability [63]. The descriptor is directly correlated with  $\text{Log } \Phi_{\Delta}$ . Porphine has the lowest value of the descriptor (3.54) while (p-I)<sub>4</sub>-TPP has the highest value of *SpMax4\_Bh(p)* equal to 4.02 (see Table 7). The diagonal elements are roughly proportional to the polarizability of the atoms. The larger the polarizability of the atoms the larger the *SpMax4\_Bh(p)* value, which leads to a increase in  $\text{Log } \Phi_{\Delta}$ .

*SI\_fosc* (relative contribution 8.4%) is an oscillator strength for the first electronic transition  $S_0 \rightarrow S_1$  calculated with M06-2X/6-31G(d,p),SDD method in the gas phase. The descriptor is directly correlated with  $\text{Log } \Phi_{\Delta}$ . It was already involved in Model 2 and was inversely correlated with  $\text{Log } \Phi_{\Delta}$ . However, in Model 5 it is directly proportional to  $\text{Log } \Phi_{\Delta}$ . It reveals that metalloporphyrins, which were added to the dataset, have a relatively high values of *SI\_fosc* (Table 7) and have high values of  $\text{Log } \Phi_{\Delta}$ . Nevertheless, the descriptor is barely correlated with  $\text{Log } \Phi_{\Delta}$ :  $R^2 = 0.0029$ .

*HL\_Gap* (relative contribution 1.2%) is a HOMO-LUMO gap calculated at B3LYP/6-31G(d,p) level of theory for the gas phase. The descriptor is directly correlated with  $\text{Log } \Phi_{\Delta}$  and has a minor contribution to the model. Several compounds including 3,7,12,18-tetramethyl-2,8-diethylporphyrin, Zn-octaethylporphyrin, and Zn-TPP possess the highest value of HOMO-LUMO gap equal to 2.95 eV (see Table 7) while tetraphenyltetrahydroporphyrin has the lowest value of HOMO-LUMO gap (2.15 eV). *HL\_Gap* descriptor reflects the fact that  $S_1$  state energy affects the rate of chemical quenching of <sup>1</sup>O<sub>2</sub> [26]. Thus, the molecules with low *HL\_Gap* value participate in <sup>1</sup>O<sub>2</sub> quenching and do not generate singlet oxygen or do it less efficiently.

Experimental value of  $\text{Log } \Phi_{\Delta}$  and values by Model 5 for each porphyrin and metalloporphyrin compound are presented in Table S5 and in Fig. 5.

We additionally splitted the dataset four times and tested the model performance in each case. The statistical parameters had following value ranges:  $R^2 = 0.891-0.936$ ,  $\text{pred}_R^2 = 0.719-0.791$ ,  $q^2 = 0.745-0.873$ , and  $\text{RMSE} = 0.030-0.047$ . Thus, statistical parameters after alternative

dataset splitting are sufficient for Model 5.  $R^2$  parameter values of Model 4 (SVR) are higher, but the key parameter  $\text{pred\_}R^2$  is higher for Model 5.

### 3.6 Model 6 (RFR)

Next, we have built a random forest regression model (Model 6). The model used 70 descriptors and 12 trees with a maximal depth equal to 80. Two descriptors had a relative contribution more than 10%. We will regard these descriptors in details.

$SpDiam\_EA(bo)$  (relative contribution -37.4%) is a spectral diameter from edge adjacency matrix weighted by bond order [59-61]. The descriptor is inversely correlated with  $\text{Log } \Phi_{\Delta}$ . It was already used in Model 3. Tetraphenyltetrahydroporphyrin has the highest value of  $SpDiam\_EA(bo)$  equal to 7.79 while porphine has the lowest value of  $SpDiam\_EA(bo)$  equal to 7.22 (Table 8). Apparently, bulky substituents are unfavorable for high quantum yield of  $^1O_2$  generation as they hinder the interaction of excited states of porphyrin and metalloporphyrin compounds with triplet state molecular oxygen that leads to reduced  $\Phi_{\Delta}$  values.

**Table 8.** Molecular descriptor values used in Model 6 (RFR).

Compound	SpDiam_EA(bo)	saaNH
(m-Cl) <sub>4</sub> -TPP	7.57	7.47
(o-Cl) <sub>4</sub> -TPP	7.58	7.43
(o-F) <sub>4</sub> -TPP	7.58	6.84
(p-Br) <sub>4</sub> -TPP <sup>test</sup>	7.56	7.62
(p-Cl) <sub>4</sub> -TPP	7.56	7.5
(p-F) <sub>4</sub> -TPP	7.56	7.15
(p-I) <sub>4</sub> -TPP	7.56	7.68
2,3,7,8-Tetraethylporphyrin	7.54	7.2
2,3,12,13-Tetraethylporphyrin	7.54	7.3
3,7,12,18-Tetramethyl-2,8-diethylporphyrin	7.53	7.23
5,10,15-NO <sub>2</sub> -etioporphyrin I	7.78	6.29
5,10-NO <sub>2</sub> -etioporphyrin I <sup>test</sup>	7.73	6.68
5,15-NO <sub>2</sub> -etioporphyrin I	7.71	6.68
5-CHO-Octaethylporphyrin	7.65	7.51
5-NH <sub>2</sub> -etioporphyrin I	7.63	7.38
5-NO <sub>2</sub> -etioporphyrin I	7.69	7.07
Deuteroporphyrin-IX-DME <sup>test</sup>	7.53	7.07
Etioporphyrin I	7.56	7.45
N-CH <sub>3</sub> -Octaethylporphyrin <sup>test</sup>	7.7	3.84
Octaethylchlorin	7.56	7.7
Octaethylporphyrin	7.57	7.66
Octamethyl-TPP	7.77	8.1
Porphine <sup>test</sup>	7.22	6.75
Tetraphenylchlorin	7.55	7.67
Tetraphenyltetrahydroporphyrin	7.79	3.83
TPP	7.56	7.61
Zn-(5,10-NO <sub>2</sub> )etioporphyrin I	7.73	6.68
Zn-(5,15-NO <sub>2</sub> )etioporphyrin I	7.71	6.68
Zn-(o-F) <sub>4</sub> -TPP	7.58	6.84
Zn-Etioporphyrin	7.56	7.45
Zn-Octaethylporphyrin <sup>test</sup>	7.57	7.66
Zn-TPP	7.56	7.61



<sup>test</sup> designates the compounds belonging to the test set

*SaaNH* (relative contribution 12.7%) is an atom-type electrotopological E-state index for aNH<sub>a</sub> groups (NH groups accompanied with two aromatic bonds). The descriptor is directly proportional to  $\text{Log } \Phi_{\Delta}$ . Octamethyl-TPP has the highest value of *SaaNH* equal to 8.1 while tetraphenyltetrahydroporphyrin has the lowest value of *SaaNH* equal to 3.83 (Table 8). E-state indices are a combination of electronic, topological, and valence state information that summarizes information related to atom types and electron accessibility that are influenced by all of the structural features of a molecule [64-66]. *SaaNH* emphasizes electronic molecular information. For instance, octamethyl-TPP possesses electron donating methyl groups and an extended  $\pi$ -conjugated system, which leads to an increase in *SaaNH* value. On the other hand, in tetraphenyltetrahydroporphyrin the conjugation is blocked by saturation of pyrrole rings, which leads to a pronounced decrease of  $\text{Log } \Phi_{\Delta}$ .

Experimental values of  $\text{Log } \Phi_{\Delta}$  and values predicted by Model 6 for each porphyrin and metalloporphyrin compound are presented in Table S6 and in Fig. 5. As one can see, the highest residual  $\text{Log } \Phi_{\Delta}$  is equal to 0.07 for Model 6, which is lower than for Model 4 (0.10) and Model 5 (0.13). We may assume that RFR model (Model 6) is superior to SVR and MLR models.

We also calculated vertical S<sub>1</sub> and T<sub>1</sub> energies of all porphyrin compounds. Surprisingly, the singlet-triplet energy gap was not found to play a role in any of the models. In fact, the determination coefficient R<sup>2</sup> between S-T gap values and  $\text{Log } \Phi_{\Delta}$  was equal to 0.02. This is consistent with our previous work on furocoumarins, which demonstrated that S-T gap does not influence the  $\text{Log } \Phi_{\text{T}}$  prediction [25].

Key statistical parameters after alternative splitting had the following value ranges, according to Model 6: R<sup>2</sup> = 0.940-0.968, pred\_R<sup>2</sup> = 0.569-0.828, q<sup>2</sup> = 0.560-0.669, and RMSE = 0.024-0.034. The statistical parameters after alternative dataset splitting are sufficient: thus, all three models possess satisfactory tolerance to alternative dataset splitting. SVR model (Model 4) possesses the highest determination coefficients towards the training set as compared to Model 5 and Model 6. However, the MLR model possesses the highest predicting ability towards the test set and the highest internal stability (q<sup>2</sup>).

Thus, we proved that QSPR method is applicable to the study of singlet oxygen generation not only for free base porphyrins, but also for metalloporphyrins. The developed models deliver reasonable results and are theoretically well interpretable.

## 4. Conclusion

In this study, a QSPR approach for predicting the efficiency of singlet oxygen generation by free base porphyrins and metalloporphyrins was developed. Three different methods of machine learning were applied: support vector regression (SVR), multiple linear regression (MLR), and random forest regression (RFR).

First, a dataset consisting of 26 free base porphyrins was studied and three models predicting  $\Phi_{\Delta}$  values were developed. All three models were found to be sufficient. The SVR model was found to be most accurate in the prediction of  $\Phi_{\Delta}$  for the training set of compounds. The MLR model showed the best performance for predicting  $\Phi_{\Delta}$  for the test set. The RFR model was found to be the most internally stable one ( $q^2 = 0.787$ ). Thus, we proved that QSPR methodology is applicable to the study of  $^1\text{O}_2$  generation by porphyrins.

Second, we examined the activity for a larger set of molecules, comprised of 26 free base porphyrins and six metalloporphyrins. All three models were found to be sufficient. The SVR model performed as the best one in predicting  $\Phi_{\Delta}$  for the training set and was the most internally stable one ( $q^2 = 0.842$ ). The RFR model showed the best performance in predicting  $\Phi_{\Delta}$  for the test set ( $\text{pred\_R}^2 = 0.875$ ). The MLR model showed the best tolerance to alternative dataset splitting. Thus, we showed that QSPR methodology is applicable to the study of  $^1\text{O}_2$  generation by coordination compounds, in particular, metalloporphyrins.

The models were built using a combination of autocorrelation, GETAWAY, topological, and quantum-chemical descriptors. SpDiam\_EA(bo), Eig02\_EA(dm), MATS1e, and S<sub>1</sub> state energy descriptors were found to be most influential. Relative contribution of the descriptors to the models was evaluated and interpreted taking into account structural and electronic effects in the molecules. MLR models provided valuable information about the descriptors involved. For example, it proves that halogenation of porphyrins increases  $\Phi_{\Delta}$ , while introduction of nitro groups has a negative influence on  $\Phi_{\Delta}$ .

Machine learning and QSPR are largely used nowadays in photochemistry. For example, they are applied for the prediction of photovoltaic properties of new materials [67,68]. When we began our studies in 2016 [24], there were almost no works dedicated to  $^1\text{O}_2$  generation analysis with QSPR, but now our studies belong to a new emerging field of knowledge [69]. The developed models can be used not only for PDT, but also for the PUVA-therapy of such diseases as psoriasis and vitiligo [25]. We believe that the QSPR models developed in our study will be useful for the screening and design of free base porphyrin and metalloporphyrin photosensitizers with high singlet oxygen generation quantum yields.

## Conflicts of Interest

There are no conflicts to declare.

## References

- [1] Lacombe S, Pigota T. Materials for selective photo-oxygenation vs. photocatalysis: preparation, properties and applications in environmental and health fields. *Catal Sci Technol* 2016;6:1571-1592
- [2] Zhao J, Wu W, Sun J, Guo S. Triplet photosensitizers: from molecular design to applications. *Chem Soc Rev* 2013;42:5323-5351.
- [3] Vicente MD, Smith KM. Syntheses and Functionalizations of Porphyrin Macrocycles. *Curr Org Synth* 2014;11:3–28.
- [4] Cannon JB. Pharmaceuticals and drug delivery aspects of heme and porphyrin therapy. *J Pharm Sci* 1993;82:435-446.
- [5] Bera K, Maiti S, Maity M, Mandal C, Maiti NC. Porphyrin–Gold Nanomaterial for Efficient Drug Delivery to Cancerous Cells. *ACS Omega* 2018;3:4602-4619.
- [6] Imran M, Ramzan M, Qureshi AK, Khan MA, Tariq M. Emerging Applications of Porphyrins and Metalloporphyrins in Biomedicine and Diagnostic Magnetic Resonance Imaging. *Biosensors (Basel)* 2018;8(4).
- [7] Dommaschk M, Gröbner J, Wellm V, Hövener JB, Riedel C, Herges R. Dendronised Ni(ii) porphyrins as photoswitchable contrast agents for MRI. *Phys Chem Chem Phys* 2019;21:24296-24299.
- [8] Khadria A, Fleischhauer J, Boczarow I, Wilkinson JD, Kohl MM, Anderson HL. Porphyrin Dyes for Nonlinear Optical Imaging of Live Cells. *iScience* 2018;4:153-163.
- [9] Senge OM, Fazekas M, Notaras EGA, Blau WJ, Zawandzka M, Locos OB, Ni Mhuircheartaigh EM. Nonlinear Optical Properties of Porphyrins. *Adv Mater* 2007;19:2737-2774.
- [10] Spikes JD, Jori G. Photodynamic Therapy of Tumours and Other Diseases Using Porphyrins. *Laser Med Sci* 1987;2:3-15.
- [11] Tournas JA, Lai J, Truitt A, Huang YC, Osann KE, Choi B, Kelly KM. Combined benzoporphyrin derivative monoacid ring photodynamic therapy and pulsed dye laser for port wine stain birthmarks. *Photodiagnosis Photodyn Ther* 2009;6:195–199.
- [12] Wan MT, Lin JY. Current evidence and applications of photodynamic therapy in dermatology. *Clin Cosmet Investig Dermatol* 2014;7:145-163.

- [13] Clarke CP, Knight SR, Daniel FJ, Seevanayagam S. Management of malignant mesothelioma by decortication and adjunct phototherapy. *Asian Cardiovasc Thorac Ann* 2006;14:206–209.
- [14] Stylli SS, Kaye AH, MacGregor L, Howes M, Rajendra P. Photodynamic therapy of high grade glioma - long term survival. *J Clin Neurosci* 2005;12:389–398.
- [15] Stylli SS, Howes M, MacGregor L, Rajendra P, Kaye AH. Photodynamic therapy of brain tumours: evaluation of porphyrin uptake versus clinical outcome. *J Clin Neurosci* 2004;11(6):584-596.
- [16] Qiu H, Kim MM, Penjweini R, Finlay JC, Busch TM, Wang T, Guo W, Cengel KA, Simone CB 2nd, Glatstein E, Zhu TC. A Comparison of Dose Metrics to Predict Local Tumor Control for Photofrin-mediated Photodynamic Therapy. *Photochem Photobiol* 2017;93(4):1115-1122.
- [17] Sun BO, Li W, Liu N. Curative effect of the recent photofrin photodynamic adjuvant treatment on young patients with advanced colorectal cancer. *Oncol Lett* 2016;11:2071–2074.
- [18] Duchi S, Sotgiu G, Lucarelli E, Ballestri M, Dozza B, Santi S, Guerrini A, Dambrosio P, Giannini S, Donati D, Ferroni C, Varchi G. Mesenchymal stem cells as delivery vehicle of porphyrin loaded nanoparticles: effective photoinduced in vitro killing of osteosarcoma. *J Control Release* 2013;168(2):225-237.
- [19] Marciel L, Teles L, Moreira B, Pacheco M, Lourenço LM, Neves MG, Tomé JP, Faustino MA, Almeida A. An effective and potentially safe blood disinfection protocol using tetrapyrrolic photosensitizers. *Future Med Chem* 2017;9(4):365-379.
- [20] Marciel L, Mesquita MQ, Ferreira R, Moreira B, Pms Neves MG, F Faustino MA, Almeida A. An efficient formulation based on cationic porphyrins to photoinactivate *Staphylococcus aureus* and *Escherichia coli*. *Future Med Chem* 2018;10(15):1821-1833.
- [21] Sobotta L, Skupin-Mrugalska P, Piskorz J, Mielcarek J. Porphyrinoid photosensitizers mediated photodynamic inactivation against bacteria. *Eur J Med Chem* 2019;175:72-106.
- [22] Singh S, Aggarwal A, Thompson S, Tome JP, Zhu X, Samaroo D, Vinodu M, Gao R, Drain CM. Synthesis and Photophysical Properties of Thioglycosylated Chlorins, Isobacteriochlorins, and Bacteriochlorins for Bioimaging and Diagnostics. *Bioconjugate Chem* 2010;21:2136-2146.
- [23] De Souza TGB, Vivas MG, Mendonça CR, Plunkett S, Filatov MA, Senge MO, De Boni L. Studying the intersystem crossing rate and triplet quantum yield of meso-substituted porphyrins by means of pulse train fluorescence technique. *J Porphyrins Phthalocyanines* 2016;20:282–291.

- [24] Buglak AA, Telegina TA, Kritsky MS. A quantitative structure-property relationship (QSPR) study of singlet oxygen generation by pteridines. *Photochem Photobiol Sci* 2016;15:801-811.
- [25] Buglak AA, Kononov AI. Triplet state generation by furocoumarins revisited: a combined QSPR/DFT approach. *New J Chem* 2018;42:14424-14432.
- [26] Schweitzer C, Schmidt R. Physical Mechanisms of Generation and Deactivation of Singlet Oxygen. *Chem Rev* 2003;103(5):1685-1757.
- [27] Braun AM, Oliveros E. Applications of singlet oxygen reactions: mechanistic and kinetic investigations. *Pure Appl Chem* 1990;62:1467-1476.
- [28] Wilkinson F, McGarvey DJ, Olea AF. Excited Triplet State Interactions with Molecular Oxygen: Influence of Charge Transfer on the Bimolecular Quenching Rate Constants and the Yields of Singlet Oxygen [ $O^*(1\Delta_g)$ ] for Substituted Naphthalenes in Various Solvents. *J Phys Chem* 1994;98:3762–3769.
- [29] Marian CM. Spin-orbit coupling and intersystem crossing in molecules. *Comput Mol Sci* 2012;2(2):187–203.
- [30] Cowan DO, Drisko RL. The Triplet State. In: Cowan DO, editor. *Elements of Organic Photochemistry*. Springer: Boston; 1976. p. 205-266.
- [31] Zhao J, Chen K, Hou Y, Che Y, Liu L, Jia D. Recent progress in heavy atom-free organic compounds showing unexpected intersystem crossing (ISC) ability. *Org Biomol Chem* 2018,16, 3692-3701
- [32] M.A. Filatov. Heavy-atom-free BODIPY photosensitizers with intersystem crossing mediated by intramolecular photoinduced electron transfer. *Org. Biomol. Chem.*, 2020, 18, 10-27.
- [33] El-Sayed MA. Spin-Orbit Coupling and the Radiationless Processes in Nitrogen Heterocyclics. *J Chem Phys* 1963;38:2834.
- [34] Englman R, Jortner J. The energy gap law for radiationless transitions in large molecules. *Mol Phys* 1970;18:145–164.
- [35] Ganzha VA, Gurinovich GP, Dzhagarov BM, Egorova GD, Sagun EI, Shul'ga AM. Influence of the molecular structure on the quenching of triplet states of porphyrins by molecular oxygen. *J Appl Spectrosc* 1989;50:402–406.
- [36] Stewart JJ. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements. *J Mol Model* 2007;13(12):1173-213.
- [37] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys Rev A* 1988;38(6):3098–3100.

- [38] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 1988;37(2):785–789.
- [39] Hay PJ, Wadt WR. Ab initio effective core potentials for molecular calculations – potentials for the transition-metal atoms Sc to Hg. *J Chem Phys* 1985;82:270-283.
- [40] Wadt WR, Hay PJ. Ab initio effective core potentials for molecular calculations – potentials for main group elements Na to Bi. *J Chem Phys* 1985;82:284-298.
- [41] Hay PJ, Wadt WR. Ab initio effective core potentials for molecular calculations – potentials for K to Au including the outermost core orbitals. *J Chem Phys* 1985;82:299-310.
- [42] Zhao Y, Truhlar DG. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06 Functionals and Twelve Other Functionals. *Theor Chem Acc* 2008;120:215-241.
- [43] Zhao Y, Truhlar DG. Density Functionals with Broad Applicability in Chemistry. *Acc Chem Res* 2008;41:157-167.
- [44] Leininger T, Nicklass A, Stoll H. The accuracy of the pseudopotential approximation. II. A comparison of various core sizes for indium pseudopotentials in calculations for spectroscopic constants of InH, InF, and InCl. *J Chem Phys* 1996;105:1052.
- [45] Frisch MJT, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant JC, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ. *Gaussian 16, Revision C.01*, Gaussian, Inc.: Wallingford CT, 2016.
- [46] Sushko I, Novotarskyi S, Körner R, Pandey AK, Rupp M, Teetz W, Brandmaier S, Abdelaziz A, Prokopenko VV, Tanchuk VY, Todeschini R, Varnek A, Marcou G, Ertl P, Potemkin V, Grishina M, Gasteiger J, Schwab C, Baskin II, Palyulin VA, Radchenko EV, Welsh WJ, Kholodovych V, Chekmarev D, Cherkasov A, Aires-de-Sousa J, Zhang QY, Bender A, Nigsch F, Patiny L, Williams A, Tkachenko V, Tetko IV. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J Comput Aided Mol Des* 2011;25(6):533-554.

- [47] Golbraikh A, Tropsha A. Beware of  $q^2$ ! *J Mol Graph Model* 2002;20(4):269-276.
- [48] Moran PA. Notes on continuous stochastic phenomena. *Biometrika* 1950;37(1-2):17-23.
- [49] Zhang XF, Xu HJ. Influence of halogenation and aggregation on photosensitizing properties of zinc phthalocyanine (ZnPC). *J Chem Soc, Faraday Trans* 1993;89:3347-3351.
- [50] Consonni V, Todeschini R, Pavan MJ. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 1. Theory of the novel 3D molecular descriptors. *Chem Inf Comp Sci* 2002;42:682-692.
- [51] Consonni V, Todeschini R, Pavan M, Gramatica P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors. 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J Chem Inf Comp Sci* 2002;42:693-705
- [52] Shao W, Wang H, He S, Shi L, Peng K, Lin Y, Zhang L, Ji L, Liu H. Photophysical properties and singlet oxygen generation of three sets of halogenated corroles. *J Phys Chem B* 2012;116(49):14228-14234.
- [53] Marin DM, Payerpaj S, Collier GS, Ortiz AL, Singh G, Jones M, Walter MG. Efficient intersystem crossing using singly halogenated carbomethoxyphenyl porphyrins measured using delayed fluorescence, chemical quenching, and singlet oxygen emission. *Phys Chem Chem Phys* 2015;17(43):29090-29096.
- [54] Robinson JW. *Atomic Spectroscopy*. 2nd ed. CRC Press: Boca-Raton; 1996.
- [55] Demtröder W. *Laser Spectroscopy: Basic Concepts and Instrumentation*. 3rd ed. Springer: Berlin; 2003.
- [56] Filatov MA, Balushev S, Ilieva IZ, Enkelmann V, Miteva T, Landfester K, Aleshchenkov SE, Cheprakov AV. Tetraaryltetraanthra[2,3]porphyrins: Synthesis, Structure, and Optical Properties. *J Org Chem* 2012;77(24):11119-11131.
- [57] Filatov MA, Heinrich E, Busko D, Ilieva IZ, Landfester K, Balushev S. Reversible oxygen addition on a triplet sensitizer molecule: protection from excited state depopulation. *Phys Chem Chem Phys* 2015;17:6501-6510.
- [58] Krasnovskiĭ AA Jr, Venediktov EA, Chernenko OM. Quenching of singlet oxygen with chlorophylls and porphyrins. *Biofizika* 1982;27(6):966-972.
- [59] Estrada E. Spectral Moments of Edge Adjacency Matrix in Molecular Graphs. 1. Definition and Applications to the Prediction of Physical Properties of Alkanes. *J Chem Inf Model* 1996;36:844-849.
- [60] Estrada E. Spectral Moments of the Edge Adjacency Matrix of Molecular Graphs. 2. Molecules Containing Heteroatoms and QSAR Applications. *Chem Inf Model* 1997;37:320-328.

- [61] Estrada E. Spectral Moments of Edge Adjacency Matrix in Molecular Graphs. 3. Molecules Containing Cycles. *Chem Inf Model* 1998;38:23-27.
- [62] Filatov MA, Balushev S, Landfester K. Protection of densely populated excited triplet state ensembles against deactivation by molecular oxygen. *Chem Soc Rev* 2016;45:4668-4689.
- [63] Todeschini R, Consonni V. *Molecular Descriptors for Chemoinformatics*. 2nd ed. Wiley: New York; 2009.
- [64] Kier LB, Hall LH. An Electrotopological State Index for Atoms in Molecules. *Pharm Res* 1990;7:801-807.
- [65] Hall LH, Kier LB. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J Chem Inf Comput Sci* 1995;35:1039-1045.
- [66] Kier LB, Hall LH. *Molecular Structure Description: The Electrotopological State*. Academic Press: London, 1999.
- [67] Sun W, Zheng Y, Yang K, Zhang Q, Shah AA, Wu Z, Sun Y, Feng L, Chen D, Zeyun Xiao, Shirong Lu, Yong Li, Kuan Sun. Machine learning-assisted molecular design and efficiency prediction for high-performance organic photovoltaic materials. *Sci Adv* 2019;5(11):eaay4275.
- [68] Sun W, Li M, Li Y, Wu Z, Sun Y, Lu S, Xiao Z, Zhao B, Sun K. The Use of Deep Learning to Fast Evaluate Organic Photovoltaic Materials. *Adv Theory Simul* 2019;2(1):1800116.
- [69] Zhao J, Zhou Y, Li C, Xie Q, Chen J, Chen G, Peijnenburg WJGM, Zhang YN, Qu J. Development of a quantitative structure-activity relationship model for mechanistic interpretation and quantum yield prediction of singlet oxygen generation from dissolved organic matter. *Sci Total Environ* 2020;712:136450.