



Article Stability Constant and Potentiometric Sensitivity of Heavy Metal–Organic Fluorescent Compound Complexes: QSPR Models for Prediction and Design of Novel Coumarin-like Ligands

Phan Thi Diem-Tran^{1,†}, Tue-Tam Ho^{2,†}, Nguyen-Van Tuan², Le-Quang Bao², Ha Tran Phuong¹, Trinh Thi Giao Chau¹, Hoang Thi Binh Minh¹, Cong-Truong Nguyen², Zulayho Smanova³, Gerardo M. Casanola-Martin⁴, Bakhtiyor Rasulev^{3,4,*}, Hai Pham-The^{2,*} and Le Canh Viet Cuong^{1,*}

- ¹ Mientrung Institute for Scientific Research, Vietnam National Museum of Nature, Vietnam Academy of Science and Technology, Hue 53000, Vietnam
 - Faculty of Pharmaceutical Chemistry and Technology, Hanoi University of Pharmacy, 13-15 Le Thanh Tong, Hoan Kiem, Hanoi 10000, Vietnam
- ³ Department of Chemistry, National University of Uzbekistan after Mirzo Ulugbek, Tashkent 100012, Uzbekistan
- ⁴ Department of Coatings and Polymeric Materials, North Dakota State University, Fargo, ND 58102, USA
- Correspondence: bakhtiyor.rasulev@ndsu.edu (B.R.); hpham.phd@gmail.com (H.P.-T.); lcvcuong@vnmn.vast.vn (L.C.V.C.)
- ⁺ These authors contributed equally to this work.

Abstract: Industrial wastewater often consists of toxic chemicals and pollutants, which are extremely harmful to the environment. Heavy metals are toxic chemicals and considered one of the major hazards to the aquatic ecosystem. Analytical techniques, such as potentiometric methods, are some of the methods to detect heavy metals in wastewaters. In this work, the quantitative structure-property relationship (QSPR) was applied using a range of machine learning techniques to predict the stability constant (log β_{ML}) and potentiometric sensitivity (PS_{ML}) of 200 ligands in complexes with the heavy metal ions Cu^{2+} , Cd^{2+} , and Pb^{2+} . In result, the log β ML models developed for four ions showed good performance with square correlation coefficients (\mathbb{R}^2) ranging from 0.80 to 1.00 for the training and 0.72 to 0.85 for the test sets. Likewise, the PSML displayed acceptable performance with an R^2 of 0.87 to 1.00 for the training and 0.73 to 0.95 for the test sets. By screening a virtual database of coumarin-like structures, several new ligands bearing the coumarin moiety were identified. Three of them, namely NEW02, NEW03, and NEW07, showed very good sensitivity and stability in the metal complexes. Subsequent quantum-chemical calculations, as well as physicochemical/toxicological profiling were performed to investigate their metal-binding ability and developability of the designed sensors. Finally, synthesis schemes are proposed to obtain these three ligands with major efficiency from simple resources. The three coumarins designed clearly demonstrated capability to be suitable as good florescent chemosensors towards heavy metals. Overall, the computational methods applied in this study showed a very good performance as useful tools for designing novel fluorescent probes and assessing their sensing abilities.

Keywords: florescent chemosensor; QSPR; quantum DFT calculation; stability constant; potentiometric sensitivity; toxicity; coumarin

1. Introduction

The aggressive increase of industrial areas without proper waste management systems along rivers has created a huge source of water pollution [1–3]. Industrial wastewater normally consists of toxic chemicals and pollutants, which are extremely harmful to people



Citation: Diem-Tran, P.T.; Ho, T.-T.; Tuan, N.-V.; Bao, L.-Q.; Phuong, H.T.; Chau, T.T.G.; Minh, H.T.B.; Nguyen, C.-T.; Smanova, Z.; Casanola-Martin, G.M.; et al. Stability Constant and Potentiometric Sensitivity of Heavy Metal–Organic Fluorescent Compound Complexes: QSPR Models for Prediction and Design of Novel Coumarin-like Ligands. *Toxics* 2023, *11*, 595. https://doi.org/ 10.3390/toxics11070595

Academic Editor: Nikiforos Alygizakis

Received: 29 April 2023 Revised: 2 July 2023 Accepted: 4 July 2023 Published: 7 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). and the environment [4]. Among all the pollutants from industry, heavy metals are considered as one of the major hazards to the aquatic ecosystem as they are poisonous and nonbiodegradable in nature, even at very low concentrations [5]. Elements that have high density and are less noxious are known as heavy metals. Generally, they are transition or basic metals with a density of $>5 \text{ g/m}^3$ [6]. They can easily migrate through the water–soil interface and the water–atmosphere interactions. There are various types of metal ions inside effluents including those of chromium (Cr(III) or Cr(IV)), manganese (Mn(II) or Mn(IV)), iron (Fe(II)), copper (Cu(II)), cadmium (Cd(II)), and lead (Pb(II)), among others, which can cause serious problems to the environment and living systems [5,7]. For example, long-term exposure to the ion Cd^{2+} with a higher concentration of 0.2 mg/kg through air, water, soil, and food leads to cancer and organ system toxicity such as the skeletal, urinary, reproductive, cardiovascular, central and peripheral nervous, and respiratory systems [8,9]. According to the World Health Organization (WHO), Pb²⁺ can attack the brain and central nervous system, causing coma, convulsions, and even death, at concentrations as low as $3.5 \,\mu\text{g/dL}$ [9]. Excessive intake of ion Cu²⁺ can cause liver or kidney damage [10]. A very recent report by Cao et al. showed a very alarming situation of water acidification and Cu^{2+} toxicity against estuarine ecosystems [11]. The above-mentioned examples clearly demonstrate an urgent demand for novel methods to detect and quantify heavy metals in the aqueous medium.

The analytical field has achieved significant advances over the last few decades in exploring the qualitative and quantitative information of pollutant elements [12,13]. Up to now, many techniques including ion-selective electrodes, reversed-phase high-performance liquid chromatography (HPLC), voltammetry, inductively coupled plasma mass spectroscopy (ICP-MS), plasma-atomic emission spectrometry (AES), and chemosensors (colorimetric, fluorometric, or fluorescent chemosensors) have been widely used in heavy metal ion detection [13]. Among them, fluorescent chemosensors have shown suitable detection specifications because of their visual simplicity, low cost, high sensitivity, good selectivity, and rapid response [14,15]. The first fluorescent chemosensor was introduced by Goppelsroder in 1867 for the determination of aluminum ions by forming a morin chelate with fluorescence-quenching effects. Since then, many fluorescent chemosensors have been developed for the detection of numerous different metal ions, which served as the precursors of analytical chemistry as widely known today [14]. In general, these sensors act as molecule receptors, which can sense and precisely interact with heavy metals and produce detectable signs due to the fluorescence intensity and/or fluorescence band shift. Designing such molecular systems for sensing metal ions through nakedeye detection is of high interest [15]. Different sensing mechanisms take place during sensing [12], such as fluorescence resonance energy transfer (FRET), intramolecular charge transfer (CT), chelation-enhanced fluorescence (CHEF), photoelectron transfer (PET), excimer formation, etc.

However, the conventional protocols for developing potential ligands able to selectively bind to the target ions require chemical synthesis and sensor membrane preparation and typically rely on the "trial and error" pathways by varying the compositions such as different solvent–plasticizers and different ratios between a ligand and an ion exchanger [16,17]. In addition, the sensitivity and selectivity parameters of the novel membranes must be examined by an appropriate potentiometric method. All these experiments require expensive, specialized, and cumbersome sample preparations and bulky laboratory equipment; thus, it is difficult to adapt them for multiple-metal-directed sensing applications or high-throughput screening of novel sensors [14,17–19].

In this sense, the quantitative structure–property relationship (QSPR) method [20–28], which is a computational approach composed by statistical models relating the structural features of ligands and their sensing properties, such as the stability constant or potentiometric sensitivity, have been employed to assist in the design and screening of fluorescence sensors [16,29–32]. Several achievements have been reported in this area, as summarized in Table 1. In 2019, Soloviev and colleagues developed several QSPR models able to predict the potentiometric sensitivity of sensors based on a variety of nitrogen-containing ligands (mainly diamides of pyridine and bipyridine acids) towards heavy metal ions (Cu^{2+} , Zn^{2+} , Cd^{2+} , and Pb^{2+}) with root-mean-squared errors of around 5 mV/dec based on the set of descriptors derived from substructural molecular fragments [29]. Later, Martynko et al. developed this concept further and aimed at Mg²⁺/Ca²⁺ potentiometric selectivity prediction using the QSPR based on the literature data [33]. While the attained precision of the model in prediction was not very high (± 0.5 logKsel), the model was able to distinguish reliably between the ligands with high-, medium-, and low-Mg²⁺ concentration selectivity. Recently, Vladimirova and colleagues explored the potential of QSPR modeling in the prediction of potentiometric selectivity for plasticized polymeric membrane sensors based on newly synthesized ligands [30]. Attempts to develop novel QSPR models effectively applied for chemical sensors are ongoing and are needed to address the main issues currently posed in the field: (i) the limited number of data to be used for training together with the weak performance of the developed models and (ii) the lack of models to predict the multiple sensing properties of fluorescent probes and application for the virtual design of potential chemosensors.

Year	Chemometric Methods	Metal Ions	Modeling Property	Chemical Structures
Kiani-Anbouhi et al. (2014) [31]	Genetic algorithm-multiple linear regression (GA-MLR); $R_{LOO}^2 = 0.86$; $R_{LGO}^2 = 0.86$; N = 29 ligands:	La ³⁺	Stability constant (logβL)	Diverse
Kiani-Anbouhi et al. (2015) [34]	GA-MLR; $R_{LOO}^2 = 0.91$; $Q_{Test}^2 = 0.94$; N = 24 ligands;	Sm ³⁺	Stability constant (logβL)	Diverse
Soloviev et al. (2019) [29]	Consensus MLR models; $Cu^{2+}: Q_{Det}^2 = 0.64$; RMSE = 8.3; $Zn^{2+}: Q_{Det}^2 = 0.78$; RMSE = 4.5; $Cd^{2+}: Q_{Det}^2 = 0.77$; RMSE = 4.8; $Pb^{2+}: Q_{Det}^2 = 0.76$; RMSE = 5.9; N = 35 ligands;	Cu ²⁺ , Zn ²⁺ , Cd ²⁺ , and Pb ²⁺	Potentiometric sensitivity (PS _{ML})	Nitrogen-containing ligands (mainly diamides of pyridine and bipyridine acids)
Martynko et al. (2020) [33]	MLR models; $Q_{Det}^2 = 0.4$; RMSE = 8.8; N = 67 ligands; $Q_{Det}^2 = 0.66$; RMSE = 5.3; N = 56 ligands (refined data);	Mg ²⁺ /Ca ²⁺	Selectivity coefficient logK(Mg ²⁺ /Ca ²⁺)	Amide ionophores
Vladimirova et al. (2022) [30]	Partial least-squares regression (PLS); Cu^{2+} : $Q_{Test}^2 = 0.66$; RMSE = 6.9; Cd^{2+} : $Q_{Test}^2 = 0.81$; RMSE = 4.2; Pb^{2+} : $Q_{Test}^2 = 0.64$; RMSE = 7.5; N = 35 ligands;	Cu ²⁺ , Cd ²⁺ , and Pb ²⁺	Potentiometric sensitivity (PS _{ML})	Nitrogen-containing ligands (mainly diamides of pyridine and bipyridine acids)
Kanahashi et al. (2022) [35]	Gaussian process regression; Best model (8 cations, 49 ligands, 2 experimental conditions): MAE = 1.31 ; $R^2_{Training} = 0.84$; N = 2706 ligands (<i>unpublished data</i>)	57 cations	Stability constant (logβL)	Diverse

In this work, for the first time, several QSPR models were developed in order to predict both the stability constant (β) and the potentiometric sensitivity (PS) of heterogeneous ionophores towards Cu²⁺, Cd²⁺, and Pb²⁺ ions simultaneously. The structure–property relationships (SPRs) extracted from the mathematical models were employed for designing novel coumarin fluorescent probes and assessing their sensing abilities. Finally, the newly designed ligands were evaluated for stability in complex by using quantum chemical calculations. The developability of the proposed chemosensors was also assessed by analyzing their toxicological profiles and effective organic synthesis routes.

2. Materials and Methods

2.1. Data Pre-Processing

The stability constant and potentiometric sensitivity database (Tables S1–S4 in the Supplementary Material) was compiled from numerous papers, which performed experiments to quantify these parameters for the three ions—cadmium, copper, and lead: Cd^{2+} , Cu^{2+} , and Pb^{2+} . In fact, research manuscripts reporting large datasets that are deposited in a publicly available database should specify where the data were deposited and provide the relevant accession numbers. Importantly, only data obtained from the same experimental condition, such as temperature and pressure, were considered as having excluded external factors affecting the model. In this study, the data of two sensors' properties were collected, including the stability constant (log β) for Cd²⁺ [36–39], Cu²⁺ [40–57], and Pb²⁺ [40,58,59] and the potentiometric sensitivity (PS_{ML}) [29,30].

The stability constant (β) shows how stable the complex formed between the metal and ligand is. The complex is more stable when this index rises [60]. The equation shown below is for the calculation of β when the ratio of the metal (*M*): ligand (*L*) is equal to 1:1.

$$M + L \Leftrightarrow ML \text{ then } \beta_{ML} = \frac{[ML]}{[M] \times [L]}$$
 (1)

In this regard, *M* and *L* are also considered as Lewis acids and bases, and the complex stability could be explained using the hard/soft acids/bases (HSABs) principle by Pearson.

Potentiometric sensitivity (PS_{ML}) is determined by the potential difference between reference and indicator electrodes, which are embedded in the individual aqueous solutions of metals in a range of concentrations [40]. It was calculated from the linear parts of the calibration curves obtained from the voltage values following each concentration of ion metal. The higher this index is, the larger the range of voltage potential exhibited. Defining this value will enable us to design more-accurate ion-selective membranes in the future. Each constant was also converted to the logarithm format for developing predictive models.

The ChemDraw 20.1.1 software was used to create the compound structures, which were then translated to the SMILES code so that the Dragon 6.0 software could calculate the molecular descriptors [61]. The data were preprocessed following the same protocols previously published [62,63]. Highly correlated descriptors ($R_{ij} > 0.95$), constant and near constant (Std. < 0.1), were removed. The data were normalized to avoid bias while training the models. The standard scale normalization of the data was carried out using the Scikit-learn package based on the mean and standard deviation as follows [64]:

$$x_{ij} = \frac{X_{ij} - \overline{X}_j}{\sqrt{\frac{\sum_{1}^{n} (X_{ij} - \overline{X}_j)^2}{n-1}}}$$
(2)

in which *n* is the number of compounds, X_j is the mean values of the *j*th descriptor, and x_{ij} and X_{ij} are the normalized and original values of the *j*th descriptor of the *i*th compound. After cleaning the data, they were randomly divided into a training set and a test set for model construction or validation following the protocol implemented in the QSARINS Version 2.2.4 software [65].

2.2. Construction of QSPR Models Using Statistical and Machine Learning Techniques

The quantitative structure–properties relationship (QSPR) models in this study were built based on multi-linear regression (MLR) algorithms to predict the value of the log β and PS_{ML} of the new compounds. In general, MLR can be represented as a linear combination of a subset of *X* variables (selected descriptor) that best describes the property *Y* with the regression coefficients α i and the intercept α_0 : $Y = \alpha_0 + \alpha_1 X_1 + ... + \alpha_n X_n$. The optimization approach based on the genetic algorithm (GA) was used for descriptor selection and to calculate the linear parameters of the MLR models [66]. The GA-MLR is based on the principles of evolution in biology, which seeks optimal models by iteratively modifying a set of randomly generated variable combinations using genetic operators, i.e., crossover, mutation, etc. The Q^2 under the influence of *K* (QUIK) rule in this algorithm enabled us to reduce the dimension of the model by setting a minimum threshold. If ($K_{XY} - K_{XX}$) < δK , this model will be excluded. K_{XY} is the correlation between the independent variables and the dependent variable; K_{XX} is total of correlation among the descriptors; δK is the threshold that the model trainer selects. This approach helped avoid multicollinearity. This entire procedure was carried out by using the QSARINS v2.2.4 software [65].

In the next steps, machine learning (ML) techniques were applied to examine the non-linear structure–property relationships (SPRs) of the chemical data. To do so, different ML algorithms implemented in the Scikit-learn library were employed using the same variable pools selected by the MLR models for each metal complex. Herein, ten algorithms, including non-linear regression (NR), the Gaussian process (GP), the decision tree (DT), support vector regression (SVR), multilayer perceptron (MLP), the *k* neighbors regressor (KNR), random forest (FR), the AdaBoost regressor MLR, the AdaBoost regressor DT, and the gradient boosting regressor were used to obtain ML models and then compared to the performance of the GA-MLR model [64]. At this stage, only the best model showing the highest performance in both the cross-validation and external validation was chosen for further screening assays. The runtime parameters corresponding to the ML models developed for each property are provided in Tables S5–S10 (Supplementary Materials). The correlation coefficient (R^2) was calculated from the experimental values on the ordinate axis, and models with $R^2 < 0.6$ were not included in this analysis [67].

2.2.1. Performance Assessment

QSPR models were evaluated through some statistical parameters. The first one is R^2 , which determines the proportion of variance in the dependent variable that can be explained by the independent variable [68] and is calculated by this equation:

$$R^{2} = 1 - \frac{\sum (y_{i} - \hat{y}_{i})^{2}}{\sum (y_{i} - \overline{y}_{i})^{2}} = \frac{\sum (\hat{y}_{i} - \overline{y}_{i})^{2}}{\sum (y_{i} - \overline{y}_{i})^{2}}$$
(3)

 y_i and \hat{y}_i are the experimental and predictive values of the dependent variable, respectively; \overline{y}_i is the average value of y_i (all of them are in the training set).

However, it should be noted that the MLR model's R^2 increased as the number of variables increased, which gradually resulted in overfitting. Therefore, the quality of the model was assessed using this adjusted parameter [68]:

$$R_{adj}^2 = 1 - (1 - R^2) \times \left(\frac{n - 1}{n - p - 1}\right) \tag{4}$$

where the *n* and *p* parameters are the number of compounds and independent variables in the training set, respectively.

In addition, we also checked for the fitness of the MLR models during the evolution process by calculating the Friedman lack-of-fit (LOF) measurement [69]. According to the analysis of Bondarchuk [70,71], this parameter is more resistant to overfitting in comparison to the commonly used least-squares error.

$$LOF = \frac{SSE}{N \left[1 - \lambda \left(\frac{c + dp}{N} \right) \right]^2}$$
(5)

In this form, the *SSE* is the sum of squares of errors; *c* is the number of terms in the model; p is the total number of descriptors; *N* is the number of data rows in the training set; λ is a safety factor; *d* is a scaled smoothing parameter [70]. Accordingly, a model with a lower LOF value is better. However, there is no universally accepted threshold to determine what constitutes a low LOF value, as this can depend on many factors such as the nature of the data, the quantity of data, and the specific algorithm used. Therefore, to determine whether this value is acceptable or not, we performed the F-test of the LOF, taking into consideration the *p*-values (<0.01) as the F-statistic for the LOF. In this sense, a *p*-value < 0.01 indicates that the LOF of selected model is significantly lower than the other models, suggesting a better-fitting model.

2.2.2. Internal Validation

 R^2 and R^2_{adj} merely demonstrate the goodness-of-fit of QSPR models. Cross-validation strategies such as leave-one-out (LOO) and leave-many-out (LMO) were applied to ensure the model's stability [72]. The result from the LOO approach is the cross-validation correlation coefficient Q^2_{LOO} , calculated by this formula:

$$Q_{LOO}^{2} = 1 - \frac{\sum (y_{i} - \hat{y}_{i/i})^{2}}{\sum (y_{i} - \overline{y}_{i})^{2}}$$
(6)

 y_i is the experimental property of the *i*th compound, and $\hat{y}_{i/i}$ is the predictive one from the model built by n - 1 remaining compounds (not including compound *i*).

LMO was also performed to overcome the disadvantages of the LOO method, which might exaggerate the predictive power of the model [73]. Then, 70% of the training set was randomly selected to construct the model for the prediction of the remaining 30%. This process was repeated 2000 times and the Q^2 value calculated each time. These values should be close to Q^2_{LOO} and to their mean (Q^2_{LMO}).

Another issue might come out that the model accidentally has the accuracy of prediction. To exclude this possibility, Y-randomization was applied by randomly rearranging the values of the dependent variables. Calculate R^2Y , Q^2Y each time 2000 times. These values should be far from the original value of the models.

2.2.3. External Validation

The test set was used to validate the predictive ability of the model by calculating the indexes below [65,73]:

$$Q_{F1}^{2} = 1 - \frac{\sum (y_{i} - \hat{y}_{i})^{2}}{\sum (y_{i} - \overline{y}_{T})^{2}}$$
(7)

$$Q_{F2}^{2} = 1 - \frac{\sum (y_{i} - \hat{y}_{i})^{2}}{\sum (y_{i} - \overline{y}_{EXT})^{2}}$$
(8)

$$Q_{F3}^{2} = 1 - \frac{\sum (y_{i} - \hat{y}_{i})^{2} / n_{EXT}}{\sum (y_{i} - \overline{y}_{Tr})^{2} / n_{Tr}}$$
(9)

where y_i and $\hat{y}_{i/i}$ are the experimental and predictive values of the property of the compound i^{th} , respectively; \overline{y}_{Ext} is average of the experimental values of the property in the training set; n_{Ext} and n_{Tr} are the number of compounds in the test set and training set, respectively.

In general, if $R^2 > 0.6$ and $Q^2 > 0.5$, the model is considered to be a good one. The closer R^2 and Q^2 are to 1, the better the obtained models are.

2.3. Applicability Domain

The QSAR/QSPR models need to determine the applicability domain (AD). AD is an area defined by the molecular descriptors and response variables of the training set. We calculated the leverage value to determine if a substance belonged to the AD following this equation [74]:

$$h_i = x_i^T (X^T X)^{-1} x_i (10)$$

 h_i is the leverage value of the *i*th compound; x_i is the molecular descriptor of the compound; *X* is matrix including *k* variables (*k* columns) of *n* compounds (*n* rows).

The AD is limited by a line of value $h^* = 3p'/n$ with p' as the number of variables of the model plus 1 and n as the number of compounds in the training set. If a compound has $h_i > h^*$ and the predicted value is outside of three-times the standard residual, it may not be predicted well by the model, or the predictive value is not reliable in this circumstance [63].

2.4. Design of Novel Coumarin-like Structures

Coumarin fluorescent probes have received significant interest in recent years due to their favorable fluorescence properties, including a high fluorescence quantum yield, long emission wavelengths, and photostability [75]. In this study, several coumarin derivatives were designed based on the precursor probes previously reported using the Marvin Sketch R-group attachment module, ChemAxon (http://www.chemaxon.com, accessed on 5 November 2022). These scaffolds showed very good selectivity in the recognition of most competing metal ions, including Cd²⁺, Cu²⁺, and Pb²⁺. By modifying the groups that have a different effect on the efficiency of charge transfer and the electron configuration of the benzene and naphthalene rings of the coumarin derivatives, it is expected that compounds will be identified with the enhancement of the fluorescence intensity towards all three metals.

2.5. Quantum Chemical Calculations

Density functional theory (DFT) calculations were performed for the ground-state gas-phase geometry optimization of the metal complexes using Beck's three-parameter exchange functional Lee–Yang–Parr (B3LYP) methods combined with the LanL2DZ basis set in the gas phase [76]. The consensus measurements were considered for estimating the stability of the complexes formed. The calculations of the newly designed compounds were performed by using the GAMESS quantum chemistry package, and the wide range of parameters were analyzed including the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) energies, among others [77]. Besides computing the HOMO–LUMO gap $\Delta E = E_{LUMO} - E_{HOMO}$, the Mulliken electronegativity χ and hardness η values could be extracted based on the finite difference approximation, viz. $\chi = (IE + EA)/2$ and $\eta = (IE - EA)/2$, where IE is the ionization energy or ionization potential and EA is the electron affinity of the ionic states of the complex system. Based on the semiquantitative approach proposed by Zhan et al., IE, EA, χ , and η can be estimated based on the calculated HOMO and LUMO energies [78]. Other properties of interest included the global softness $\sigma = 1/\eta$, which measures the extent of chemical reactivity, and the chemical potential of the complex system (μ), which is the first derivative of the total energy with respect to the number of electrons [79]. Parr and colleagues proposed a new global chemical reactivity parameter, namely an electrophilicity index (ω), which may be calculated via the equation $\omega = \mu^2/2\eta$ [80]. In order to take into account the diffuse characteristics of the molecular orbitals of heavy elements with a large electron density, all of these derived parameters were made at the B3LYP/LANL2DZ level.

2.6. Physicochemical and Toxicological Profiling

Given the broad application of the designed coumarin fluorescent probes, especially in an aqueous medium, it is of pivotal importance to determine the possible effects of the chemosensors for human and aquatic organism health. Important physicochemical properties mainly governing the aqueous solubility, organism permeability, and in vivo dermal and oral bioavailability were calculated using consensus cheminformatics approaches such as SwissADME [81] and ADMETlab 2.0 [82], among others [83–86]. To do so, chemical SMILES codes were used as the input of the computational tools, and the results were extracted based on more than 100 indices calculated previously. In addition, multiple toxicological endpoints of the newly developed sensors towards aquatic organisms were also investigated using the same tools [81,82].

2.7. Theoretical Proposal of Synthetic Accessibility for Promising Candidates

The prediction of the synthetic accessibility (SA) of the potential probe candidates was carried out using the GASA [87], ADMETlab 2.0 [82], and SwissADME tools [81]. GASA employs a rapid graph-based molecular synthetic accessibility prediction method, which has been demonstrated to be comparable to the existing methods, such as SYBA, SCScore, RAscore, and SAscore, among others. Based on the probability of the predictions,

compounds can be classified as easy- (ES) or hard-to-synthesize (HS) structures [87]. This model was trained on 640,000 compounds and achieved an accuracy of 0.932 with an AUC of 0.984. In ADMETlab 2.0, the SA scores can be predicted with ~90% accuracy using the fragment-based approach developed by Ertl and Schuffenhauer [88], on the basis of 934,046 representative molecules from the PubChem database. In addition, the SwissADME tool allows calculating the SA scores based on a fragmental contribution approach. This method consists of the analysis of nearly 13 million molecules with the hypothesis that the more frequent a chemical fragment, the easier it is to synthesize that molecule [81]. The SA scores computed by SwissADME receive a value ranging from 1 (very easy synthesis) to 10 (very difficult synthesis).

3. Results and Discussion

3.1. Development and Selection of GA-MLR QSPR Models

A total number of six MLR models were constructed based on the experimental log β and PS data. Applying an extensive feature selection process of >105 iterations by the GA, 90% of the new generations showed the same fitness results. The best MLR models were selected from the final iteration taking into account Austin's criteria (5 × number of variables \leq number of observations) to avoid the overfitting problem [66,89]. The best models were further analyzed and compared based on the quality of the statistical parameters, as shown in Table 2. The mathematical formulas of the six models are given by Equations (10) to (15), and more details about the variables selected can be found in Table S11, Supplementary Material.

$\begin{split} Log \beta_{CdL} &= 9.2894 - 2.6191 \times RBN + 6.0044 \times SPI + 15.7303 \times H_D/Dt \\ &- 4.6761 \times MATS1m - 11.3909 \times MATS2m + 4.987 \times MATS6i \end{split}$	(11)
$\begin{split} Log \beta_{CuL} &= 8.2894 + 7.1518 \times nHM + 5.5297 \times ATSC4e + 5.6081 \times MATS1v \\ &+ 9.7064 \times MATS6p - 10.5641 \times MATS7p + 7.4692 \times GATS5p \\ &+ 12.2912 \times P_VSA_MR_7 - 19.6671 \times SM07_AEA(ed) \\ &+ 8.5756 \times O - 057 + 12.5774 \times NssCH2 - 1.8575 \times B06[C - O]? \end{split}$	(12)
$\begin{aligned} Log \beta_{PbL} &= 4.6561 - 1.1628 \times nR05 - 6.1855 \times J_Dt + 6.4076 \times ATSC1s \\ &+ 6.4211 \times GATS8m - 6.0816 \times SM03_EA(bo) \end{aligned}$	(13)
$\begin{split} PS_{CdL} &= -\ 32.5684 - 13.696 \times J_B(s) + 16.1241 \times GATS8i + 27.1608 \times JGI8 \\ &+\ 30.3906 \times SpMax4_Bh(m) - 21.3087 \times SpMin3_Bh(s) \\ &+\ 38.7644 \times SpMin5_Bh(s) + 9.2121 \times DLS_01? \end{split}$	(14)
$\begin{split} PS_{CuL} &= 35.4214 - 16.7676 \times Ho_Dz(p) - 12.1368 \times GGI10 \\ &- 29.2904 \times P_VSA_p_2 + 23.0366 \times Chi1_EA(dm) \\ &+ 17.6022 \times DLS_04 - 42.4502 \times LLS_01? \end{split}$	(15)
$\begin{split} PS_{PbL} &= -53.7619 + 16.311 \times MATS6v + 30.6395 \times MATS2i \\ &- 10.6447 \times GATS4i - 50.3016 \times GGI8 + 45.0039 \times JGI8 \\ &+ 71.9904 \times SpMin5_Bh(s) + 24.299 \times Eta_F_A? \end{split}$	(16)

In general, all six models presented a goodness-of-fit with an R^2 and R^2_{adj} of approximately 0.9, except for the $log\beta_{PbL}$ -only models, whose values were ~0.8 (Figure 1). According to the Friedman LOF parameters, the $log\beta$ model of the Cu complex showed the highest value of 7.05, and the lowest value corresponded to the $log\beta$ model of Pb (1.35). As can be seen in Table 2, the F-statistics for the F-test of the LOF were highly significant (*p*-value ranging from 1.095×10^{-19} to 3.087×10^{-7}), suggesting an acceptable level of the LOF. From this test, we can confirm that no overfitting existed in the models, as they presented a good fit with a minimum number of descriptors. Considering the internal validation, the values of Q^2_{LOO} for almost all models were greater than 0.8 with the model

MLR-log β_{PbL} being the lone exception, owning an approximate Q^2 value of 0.7. For the LMO approach, the Q_{LMO}^2 values after each iteration showed a slight change, being around the Q_{LOO}^2 of the primary model (Mod. Q^2) (see Figure S1). This indicated that the Q_{LOO}^2 index in these models does not overate their predictive ability. In addition to the above criteria, the optimal model was chosen based on the lower level of $R^2 - Q_{LOO}^2$, the correlation among the independent variables (as low as possible), and the high correlation between the dependent variables and independent variable (high δK values). Alternatively, Y randomization also demonstrated that the accuracy of these models was not an accidental correlation. We found that R^2 , Q^2 , and K_{XY} in each iteration were less than the original model (Figure S2). The autocorrelation between the descriptors was checked through the Pearson correlation matrix and the correlation among the descriptors (K_{XX}). As the results show in Figures S3–S8, no significant correlation ranged from 0.25–0.48. All the descriptor values calculated by the Dragon software for each of the training and test set compounds included in the six databases are provided in Tables S21–S26.

Table 2. Overview of the performance parameters of the six MLR models.

	Statistical Parameters ¹	MLR-log _{βCdL}	$MLR-log\beta_{CuL}$	MLR-log _{βpbl}	MLR-PS _{CdL}	MLR-PS _{CuL}	MLR-PS _{PbL}
Fitting criteria	K_{XX} δK R^2 R^2_{adj} LOF p-value (F-test)	$\begin{array}{c} 0.4121\\ 0.0698\\ 0.9257\\ 0.9079\\ 2.6866\\ 6.522\times10^{-13} \end{array}$	$\begin{array}{c} 0.2649\\ 0.0577\\ 0.9085\\ 0.8862\\ 7.0495\\ 1.095\times10^{-19} \end{array}$	$\begin{array}{c} 0.2384\\ 0.0679\\ 0.8084\\ 0.7648\\ 1.3524\\ 3.087\times 10^{-7} \end{array}$	$\begin{array}{c} 0.4845\\ 0.0193\\ 0.9160\\ 0.8957\\ 2.4842\\ 6.602\times 10^{-14} \end{array}$	$\begin{array}{c} 0.3958\\ 0.0679\\ 0.8904\\ 0.8669\\ 3.1127\\ 3.491\times 10^{-12} \end{array}$	$\begin{array}{c} 0.2546 \\ 0.0554 \\ 0.9264 \\ 0.9080 \\ 3.7405 \\ 3.399 \times 10^{-14} \end{array}$
Internal validation criteria	$\begin{array}{c} Q_{LOO}^2 \\ R^2 - Q_{LOO}^2 \\ Q_{LMQ}^2 \\ Y_{scr} R^2 \\ Y_{scr} Q^2 \end{array}$	0.8637 0.0620 0.8311 0.1935 -0.3624	0.8434 0.0651 0.8209 0.1975 -0.3143	0.6918 0.1165 0.5872 0.1869 -0.5988	0.8760 0.0400 0.8556 0.1938 -0.3579	0.8423 0.0481 0.8246 0.1757 -0.3038	0.8668 0.0596 0.8394 0.1961 -0.4169
External validation criteria	$Q_{F1}^2 \ Q_{F2}^2 \ Q_{F3}^2$	0.8324 0.8120 0.9283	0.8326 0.8276 0.8415	0.6357 0.6305 0.6961	0.8791 0.8789 0.9305	0.8914 0.8793 0.8080	0.7079 0.7054 0.7734

¹ The statistical parameters correspond to the best MLR obtained based on the GA.

In the external validation, the Q_{F1}^2 , Q_{F2}^2 , and Q_{F3}^2 of the six models were almost higher than 0.8, except the Pb²⁺ models with Q_{F1}^2 values between 0.6 and 0.7, which are still acceptable according to our criteria (Table 2). From the results, all the MLR models were able to predict the log β and PS_{ML} values of the unknown structures with high accuracy.

The applicability domain of each model is shown in Figure 2. If a new chemical is to be predicted, it must fall within the AD of the model. The outliers of the domain included two types: *Y* outlier and *X* outlier. The *Y* outlier indicates that the predictive value is not unreliable, and the *X* outlier shows that these compounds should not be applied to the model. There were no *Y* outliers in all six plots, so the quality of these models was ensured. Regarding the *X* outlier, there were a few compounds that are outside of the domain. Therefore, we need to pay more attention to these structures [63].

3.2. Machine Learning Algorithms for Improving Model Performance

As can be revealed from the validation of the MLR models, the statistical linear algorithm still showed weak performances; for example, the models to predict the sensing properties of the Pb²⁺ complexes only had a testing Q^2 of 0.64–0.71 (Table 2). To overcome this challenge, machine learning (ML) methods were applied using the same variable subset selected by the GA-MLR. A wide range of 10 ML algorithms were explored adopting the Scikit-learn package in Python using the same script we developed previously [64]. To optimize the model performance, a parameter tuning search was applied, and the entire results are listed in Tables S5–S10, the Supplementary Materials. Note that only models

showing an accuracy higher than 0.6 were retained for this analysis, and the models with the highest performance were used for further virtual screening assays. Figure 3 displays the comparison histogram of every ML model against the best MLR models obtained previously. In general, there was a large dissimilarity among the modeling techniques, and the ML algorithms normally outperformed the simple linear models in both *k*-fold cross-validation on the training set and the external test set. A detailed comparison of all the selected regressors is shown in Tables S12–S17, and the resulting runtime parameters of the selected models are given in Table 3.



Figure 1. Scatter plots illustrating the correlation between the experimental and predictive values of the six QSPR models.



Figure 2. William plot showing the applicability domain of each model.

1.00

0.95

Squared Correlation Coefficient

0.65

0.60

1.00

0.95

Squared Correlation Coefficient

0.65

0.60

1.00

0.95

Squared Correlation Coefficient

0.65

0.60

Gradient Boosting Regressor AdaBoost Regressor DT MLP Regressor The best MLR model Nonlinear Regression

AdaBoost AdaBoost Regressor Regressor DT MLR

Nonlinear Regression AdaBoost Regressor MLR The best MLR model

ML-Log β_{CuL}

Nonlinear MLP Regression Regressor

ML-Log β_{PbL}

Gaussian Process Regressor

Support Vector Regresso

The best MLR model

Training Set

Training Set

ML-Log β_{CdL}

Gaussian Process Regressor Random Forest Regressor

Training Set Test Set 1.00

0.95

Squared Correlation Coefficient

0.65

0.60







Nonlinear Regression AdaBoost Regressor MLR Gaussian Process Regressor

The best MLR model Random Forest Regressor

Data	Selected Model and Method	Optimized Parameter Settings ¹
logβ (Cd ²⁺ complex)	Model M1/AdaBoost Regressor MLR	Base_estimator=LinearRegression; copy_X=True, fit_intercept=True, n_jobs=None, normalize='deprecated', positive=False); learning_rate=1.0; loss='linear'; n_estimators=50; random_state=None; fitting 5 folds for each of 100 candidates; totaling 500 fits {'n_estimators': 4, 'loss': 'square', 'learning_rate': 0.01} Alpha=0.9; ccp_alpha=0.0; criterion='friedman_mse'; init=None;
PS _{ML} (Cd ²⁺ complex)	Model M2/Gradient Boosting Regressor	<pre>learning_rate=0.1; loss='squared_error'; max_depth=3; max_features=None; max_leaf_nodes=None; min_impurity_decrease=0.0; min_samples_leaf=1; min_samples_split=2; min_weight_fraction_leaf=0.0; n_estimators=100; n_iter_no_change=None; random_state=None; subsample=1.0, tol=0.0001; validation_fraction=0.1; verbose=0, warm_start=False {'n_estimators': 91, 'min_samples_split': 10, 'min_samples_leaf': 1, 'may_foatures'('cart' 'may_denth' & 'laorming_rate('0.5)</pre>
$log\beta$ (Cu ²⁺ complex)	Model M3/GA-MLR	Equation (11)
PS _{ML} (Cu ²⁺ complex)	Model M4/Gradient Boosting Regressor	C=1.0, cache_size=200, coef0=0.0, degree=3, epsilon=0.1, gamma='scale', kernel='rbf', max_iter=-1, shrinking=True, tol=0.001, verbose=False
logβ (Pb ²⁺ complex)	Model M5/Gradient Boosting Regressor	<pre>{'kernel': 'linear', 'gamma': 10.0, 'epsilon': 0.5, 'C': 100.0} Alpha=0.9; ccp_alpha=0.0; criterion='friedman_mse'; init=None, learning_rate=0.1, loss='squared_error'; max_depth=3; max_features=None; max_leaf_nodes=None; min_impurity_decrease=0.0; min_samples_leaf=1; min_samples_split=2; min_weight_fraction_leaf=0.0; n_estimators=100; n_iter_no_change=None; random_state=None; subsample=1.0; tol=0.0001; validation_fraction=0.1; verbose=0, warm_start=False {'n_estimators': 57, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'sqrt', 'max_depth': 8, 'learning_rate': 1}</pre>
PS _{ML} (Pb ²⁺ complex)	Model M6/AdaBoost Regressor MLR	Base_estimator=LinearRegression; copy_X=True; fit_intercept=True; n_jobs=None; normalize='deprecated'; positive=False; learning_rate=1.0; loss='linear'; n_estimators=50; random_state=None; fitting 5 folds for each of 100 candidates; totaling 500 fits {'n_estimators': 42, 'loss': 'linear', 'learning_rate': 0.1}

Table 3. Resulting hyperparameters for the best QSPR models selected for predicting the log β and PS_{ML} data.

¹ The best hyperparameter combinations are shown within curly brackets.

Regarding the models to predict the stability constant $(\log\beta)$ values of the Cd²⁺–ligand complexes, the histogram shown in Figure 3 indicated that the accuracy of the AdaBoost regressor MLR and non-linear regression were comparable and higher than other models, being 0.922/0.820 and 0.926/0.812 for the training/testing R^2 , respectively. These results were slightly higher than the best MLR model, especially the training process, which resulted in an R^2 value of 0.908. AdaBoost has been described as an ensemble boosting algorithm, which repeatedly combines regressors, i.e., MLR or decision trees (DTs), on the same training set with a special focus on the error of the current prediction and the difficult cases. Meanwhile, the non-linear regression algorithm, unlike the statistical linear estimator, allows for a more complex relationship between the dependent and independent variables [90]. Herein, three functions were tested, including the polynomial, logarithmic, and exponential regressions. Given the importance of external testing, the AdaBoost regressor MLR was selected for further use in the new case prediction.

For the potentiometric sensitivity (PS_{ML}) data of the Cd²⁺–ligand complexes, seven ML models are shown, the gradient boosting, random forest, and AdaBoost DT being

the regressors with the highest accuracy for the training set ($R^2 \sim 1.0$). However, only gradient boosting and the AdaBoost DT significantly outperformed the other models with a testing R^2 of 0.948 and 0.941, respectively. Unlike AdaBoost, gradient boosting uses the gradient decent algorithm to add the weak learners in a stagewise manner so that the overall prediction error significantly decreases. Consequently, the gradient boosting model was selected for the PS_{ML} value prediction of the Cd²⁺–ligand complexes.

In contrast to the Cd²⁺ complex data, most of the ML techniques showed good performance on the Cu²⁺ log β and PS_{ML} data. For the log β data, the AdaBoost regression DT significantly outperformed the other models on the training ($R^2 = 0.984$), but only obtained an accuracy of 0.721 on the test set, which was much lower than the MLR model. In this case, MLR, even having a slightly lower coefficient for the training set ($R^2 = 0.89$) in comparison with the other ML models (training $R^2 = 0.90-0.91$), surprisingly exhibited the highest accuracy on the test set. We then selected the linear model MLR for further prediction of the Cu²⁺ log β data.

The results of the ML models in predicting the PS data of the Cu²⁺–ligand complexes were somewhat the same situation as the Cu²⁺ log β data. However, we identified support vector regression (SVR) as the best model, owning an accuracy of 0.884 and 0.901 for the training and test sets. In this case, MLR was the model with the lowest performance on both the training (training $R^2 = 0.867$) and test sets (average testing $R^2 = 0.86$). More details on the SVR algorithm implemented in the Scikit-learn package can be found in the literature [64]. This model was subsequently used in the next screening assays based on the Cu²⁺ PS_{ML} data.

The accuracy results from the Pb²⁺ log β and PS_{ML} data were very encouraging. We found significant improvement when using ML techniques in comparison with the MLR model. For the Pb²⁺ log β data, the gradient boosting regressor once again showed the highest performance, owing to the determinant coefficient for the training and test sets of 1.0 and 0.847, respectively. Another ML algorithm that also performed well on these data was multi-layer perceptron (MLP) with a training and testing accuracy of 0.903 and 0.853, respectively; however, the training results were lower than gradient boosting. Consequently, the gradient boosting model was without a doubt selected for further prediction of the Pb²⁺ log β data.

For the Pb^{2+} PS_{ML} data, the AdaBoost regressor MLR displayed high performance with the accuracy on the training and testing sets being 0.922 and 0.816, respectively. One these data, the MLR model also performed well on the training set with an R^2 value of 0.908, but the accuracy on the test set significantly dropped to ~0.7, resulting in the worst model in the external validation. Thus, the AdaBoost regressor MLR was selected for predicting the PS_{ML} values of the Pb²⁺–new ligand complexes.

3.3. Design of Coumarin-like Chemical Library for Screening of Novel Chemosensors

One of the main applications of QSPR/QSAR approaches is to screen novel active ligands from a large chemical library of a determined scaffold. Bearing in mind that coumarin fluorescence probes for heavy metal ions currently receive increasing interest due to their highly variable size, hydrophobicity, and chelation, in this study, we focused on the design and screening of novel fluorescent probes bearing coumarin moieties [75,91]. In addition, Schiff bases are attracting more and more attention in analytical fields as they provide an electron-rich environment that is favorable to binding with metal ions [92]. When combined with a fluorescence moiety such as coumarin structure, the functional groups can participate in several reactions after being in contact with metal ions so that the Schiff base ligand provides detectable signals such as emission.

In 2017, Fan reported the synthesis of several coumarin-derived Schiff bases, which showed significant fluorescence enhancement owing to inhibiting the photo-induced electron transfer (PET) process [93]. The structures were mainly based on condensing 7-amino-4-methyl-coumarin (AMC) with salicylic aldehyde and 2-hydroxy-1-naphthaldehyde. Fan also evaluated the selectivity of his derivatives upon the addition of several metal ions

such as Cd²⁺ and Pb²⁺, among others (15 metals). The fluorescent property of these Schiff bases appeared to be dependent on the charge transfer and the electron configuration of the benzene and naphthalene rings. However, Cu²⁺ was not tested for the coumarins designed in Fan's study [93]; thus, this core structure was explored in this study.

Going through the literature, we encountered coumarin-pyridin hybrids reported by Jung et al., which showed promising fluorescence-quenching properties in both the femtosecond time-resolved fluorescence (TRF) upconversion technique and ab initio calculations [47]. The authors identified that the n-picolylamide moiety (n = 1-3) allowed efficient tridentate complexation for Cu²⁺ preferred over a variety of other common heavy and toxic metal ions. Of interest is also the chemosensor bearing acetyl-7-hydroxycoumarin moiety reported by Chang and colleagues, which exhibited a highly selective and sensitive fluorescence-sensing ability for Cu⁻⁺ over other metal ions under physiological conditions [94]. We also consulted the most recent advances in the field from literature reviews [75,91].

Taking all the above-mentioned findings, here, we tried to construct a small library of coumarin derivatives by using a structure-resembling approach using the R-group attachment module implemented int the Marvin Sketch package. As a result, over 129 coumarin structures were created, as shown in Figure 4A,B. All newly designed substances were checked to see if they lied in the range of the AD. Their calculated leverage values (\hat{h}_i) were all smaller than the threshold value (h^*) of all the models, which means that QSPR models could work correctly on the chemical space of the designed compounds as they fell into the AD of the models.



Figure 4. Design and screening of the virtual chemical library of coumarin derivatives as possible fluorescent probes. Chemical design was based on: (**A**) coumarin-derived Schiff bases, and (**B**) other core coumarin structures. Coumarins predicted as potential florescent probes are highlighted (green dots inside the circles) in comparison with the other compounds (orange dots).

3.4. Virtual Screening and Complexation Potential Predictions Using DFT Calculations

In the next step, the newly designed library was screened for identifying coumarin-like florescent probes against each heavy metal using the QSPR models. The screening results are presented in Table S18 (Supplementary Materials) in which the designed compounds are denoted as NEW, and the log β and PS values of each were estimated using the respectively selected models. Of note, a potential probe must have a high log β and a low PS, which in turn can also be translated into a high PS_{ML}⁻¹. Figure 4 shows the trade-off of the log β and PS_{ML}⁻¹ predictions, in which compounds closer to the top-right corner indicate better candidates. According to this criterion, we easily identified eight compounds, *viz.* NEW02, NEW03, NEW07, NEW21, NEW26, NEW34, NEW51, and NEW80, as potential florescent probes. Their log β and PS values estimated by all the QSPR models are shown in Table 4, and the chemical SMILES of these eight coumarins are listed in Table S19. Especially, compounds NEW02, NEW03, and NEW07 were consistently predicted to be good ligand candidates in complex with all three heavy metal ions.

Table 4. Overview of the performance parameters of the six machine learning models.

Compound ID ——	Stal	bility Constant (lo	gβ)	Potentiomet	ric Sensitivity (PS	_{ML} , mV/dec)
	Cd ²⁺	Cu ²⁺	Pb ²⁺	Cd ²⁺	Cu ²⁺	Pb ²⁺
NEW02	9.537	9.022	3.133	12.146	13.809	16.135
NEW03	10.263	9.610	2.935	14.901	16.194	16.947
NEW07	8.278	8.870	4.304	4.516	5.518	17.973
NEW21	6.912	5.506	4.024	4.746	22.307	16.812
NEW26	7.087	7.701	2.539	7.230	34.360	27.286
NEW34	6.550	6.607	3.435	4.236	28.975	23.507
NEW51	9.566	9.252	2.608	8.119	30.902	21.520

Resulting from the prediction of the complexing characteristics, three florescent probe candidates were revealed, and they were subsequently examined for their interaction with Cd^{2+} , Cu^{2+} , and Pb^{2+} using DFT calculations with B3LYP exchange functionals. To simplify the calculations, the stoichiometry of the complexes was adopted to be in a 1:1 (M:L) metal-and-coumarin-ligand molar ratio. In accordance with the literature, given the availability of a lone pair on the nitrogen atom, the Schiff base could form mono complexes with many metals; meanwhile, adding other substitutions, such as OH and SH, to the adjacent heterocycles may result in the formation of bidentate or tridentate chelates. This fact could be revealed by observing the partial distribution of the charge on the electrostatic potential surface (ESP) shown in Figure 5. This electrostatic potential map for all three compounds clearly revealed hydrophilic regions, including negative (red) and positive (blue) potentials, and hydrophobic regions, which are normally neutral (mainly the green area). The optimized geometric structures of the coumarin compounds computed via the B3LYP method showed that the reactive sites corresponding to the negative potentials were mainly localized on the O and S atoms, and two compounds, NEW02 and NEW07, showed higher polarity compared to NEW03. The total electron density of the studied derivatives varied from -0.31 to +0.29.

The quantum chemical parameters can provide insight into the chemical nature of the reaction mechanisms [95]. Through the DFT calculations at the B3LYP/LanL2DZ levels of theory, numerous descriptors corresponding to six metal complexes were determined as summarized in Table 5. Figure 6 shows the optimized structures of the metal–coumarin complexes and the corresponding distances of the π contacts between the ligands and metals. The molecular frontier orbitals (HOMO–LUMO) consisted of important descriptors mainly associated with the molecular reactivity.



Figure 5. Structures of 3 novel coumarin-like compounds that have the potential to be heavy metal binders.

Table 5. The quantum chemical parameters of three coumarin compounds calculated at B3LYP/LanL2DZ levels.

				DF	T Energy (eV)				
Global Reactivity Index	Cd	Cd ²⁺ Complexes			Cu ²⁺ Complexes			Pb ²⁺ Complexes		
	NEW02	NEW03	NEW07	NEW02	NEW03	NEW07	NEW02	NEW03	NEW07	
HOMO–LUMO Gap, ∆E	4.004	3.839	4.328	4.586	4.419	4.439	3.983	3.839	4.300	
E _{HOMO}	-6.074	-5.947	-6.305	-6.950	-6.774	-6.803	-6.225	-6.044	-6.410	
E _{LUMO}	-2.070	-2.108	-1.977	-2.363	-2.354	-2.364	-2.242	-2.205	-2.109	
Ionization Energy, IE	6.074	5.947	6.305	6.950	6.774	6.803	6.225	6.044	6.410	
Electron Affinity, EA	2.070	2.108	1.977	2.363	2.355	2.364	2.242	2.205	2.109	
Global Hardness, η	2.002	1.919	2.164	2.293	2.210	2.219	1.992	1.920	2.150	
Global Softness, σ	0.500	0.521	0.462	0.436	0.453	0.451	0.502	0.521	0.465	
Electronegativity, χ	4.072	4.027	4.141	4.656	4.564	4.584	4.234	4.125	4.259	
Electrophilicity, ω	4.142	4.225	3.962	4.727	4.715	4.734	4.500	4.431	4.219	

The E_{HOMO} descriptor is related to the ability of the molecules to donate electrons to empty molecular orbitals with low energy, and the E_{LUMO} descriptor indicates the electron-accepting affinity [96]. The interacting parts of the molecules were also identified by looking at the HOMO and LUMO visualization (Figure 7). Their different HOMO–LUMO ΔE represent the hard–soft, acid–base characteristics of the molecule. In such a way, the low gap energies identified in this study reflected the high chemical reactivity, optical polarizability, and low kinetic stability of the three metal complexes [97]. In other words, a molecule with a higher HOMO and a lower LUMO energy level has a higher chemical reactivity. Thus, for the Cd²⁺ complexes, the energy gap followed the order: NEW03 < NEW02 < NEW07 (see Table 5), indicating that NEW03 had the lowest reactivity (the most-chemically stable) followed by NEW03 and NEW07 (the most-reactive). Likewise, the HOMO–LUMO gap for the Cu²⁺ complexes were ranked as NEW03 < NEW07 < NEW02, and those for the Pb²⁺ complexes were similar to the Cd²⁺ complexes. In addition, the electron affinity (EA) and electronegativity (χ) of all six complexes were quite low compared to the ionization energy (IE). The current molecular orbital analysis also showed that the



level of the frontier orbitals and low-energy electronic transitions were almost similar, suggesting the low energy intra-ligand charge transfer states of the metal complexes [95,97].

Figure 6. Optimized structures of 3 potential coumarin-like compounds in complex with 3 heavy metal ions.



Figure 7. Frontier molecular orbitals of three complexes relevant to fluorescence quenching.

Another descriptor of pivotal interest is the chemical hardness, η , which refers to the resistance of a chemical species to the polarization of the electron cloud. It gives information, along with the electronegativity of the chemical reactivity and stability [79,95,96]. The softness is the reciprocal value of hardness. Previous studies demonstrated that most hard species have an electrostatic interaction, and soft species mainly interact through covalent bonding, e.g., mixing of orbitals. Of note, the global softness value increased from NEW07 to NEW03, while the chemical hardness and HOMO–LUMO gap values decreased. Evidently, the Schiff bases gained softer characteristics from NEW03 to NEW07. It was revealed that the stability parameters obtained in the QSPR predictions and the rankings obtained in the quantum chemical calculations were in concordance (Tables 4 and 5). In addition, Cu^{2+} and Cd²⁺ had higher electron donor characteristics (α) compared to Pb²⁺, while Cu²⁺ and Pb²⁺ showed higher basicity characteristics (β) compared to Cd²⁺. The softness parameter σ can also be calculated via equation $\sigma = \alpha/(\alpha + \beta)$, which follows the ranking Cd²⁺ > Cu²⁺ > Pb^{2+} ($\sigma = 0.96 > 0.89 > 0.85$, respectively). Thus, according to the Pearson classification for Lewis acids, Cu^{2+} and Pb^{2+} fall within the hard/soft borderline and Cd^{2+} is the softest ion. As soft acids prefer to bind to soft bases, compound NEW03, therefore, can easily interact with the softer Cd²⁺ charged ions, but not with the harder Pb²⁺ ion. As soft metals normally have a filled d subshell, according to Lancashire (https://chem.libretexts.org/, assessed on 3 March 2023), the π interactions play an important role in metal-to-ligand bonding. Complexes of three coumarins with softer metals are more stable based on the electrostatic interactions [95,97].

3.5. Physicochemical and Toxicological Profiling

Given the broad applications in the analytic field, a well-organized chemosensor must be non-toxic to living organisms when it is applied in aqueous environments [15]. Therefore, in this study, the toxicological profiles of the newly designed compounds were assessed using several cheminformatics tools, including SwissADME and ADMETlab 2.0 [81,82].

First, the physicochemical properties that may affect the interactions between chemicals and living organisms were computed for NEW02, NEW03, and NEW07. The radar chart shown in Figure 8 illustrates the multivariate data of all three coumarins compared to the physicochemical profiles of the drug-like chemical space [82]. Being small molecules (MW range from 295.3–327.5), most of the physicochemical properties of these compounds fell into the drug-like chemical space, except the lipophilicity indices (logP and logD). This means that these compounds may have a drug-like behavior including appropriate pharmacokinetic, as well as low toxicological characteristics in the human body [98]. Based on the consensus predictions of SwissADME (Table S20), the solubility of NEW02 and NEW03 was predicted to be low (logS < -6.0 mol/L), while NEW07 had higher solubility (logS = -5.4 mol/L). According to the combination of the three properties (MW, PSA, and logD_{7.4}) as established in 3PRule [86], these coumarins may transport across the biological membrane via passive diffusion with moderate-to-high permeability. The oral bioavailability of these compounds was predicted to be <20%, and the brain disposition of these chemicals appeared to be low due to their low BBB passage [83].

Subsequently, the toxic effects on the human body were forecasted using both cheminformatics and toxicophore approaches. The results shown in Table 6 suggested that all three coumarins have no noticeable effects during exposure to the human body, especially on the skin. The oral acute toxicity was quite low, but for eye contact, they all can cause irritation [75,82]. In addition, biodegradability has been considered fundamental to the assessment of the environmental exposure and risk of chemical products. As can be seen in Table 6, there was no alert fragment in the structures of three coumarins, meaning that all the chemicals are biodegradable and will have no long-term toxic effect on the environment [82].



Figure 8. Physicochemical profiles of the 3 designed coumarin-like ligands. MW, molecular weight (kDa); log P, log of octanol partition coefficient; log D, log of octanol partition coefficient at physiological pH 7.4; fChar: formal charge; log S, log of aqueous solubility (mol/L); nHD, number of hydrogen bond donors; nHA, number of hydrogen bond acceptors; nHet, number of heteroatoms; nRot, number of rotatable bonds; nRing, number of rigid bonds; nRing, number of rings; MaxRing, number of atoms in the biggest ring; TPSA, topological polar surface area (Å²).

Table 6. Relevant organo-kinetic and toxicological parameters of designed coumarins.

Parameters	NEW02	NEW03	NEW07	Description
Oral bioavailability ($F_{20\%}$)	0.982	0.999	0.968	The probability of oral bioavailability of 20% (F20%+). Category 1: $F_{20\%}$ + (bioavailability < 20%); Category 0: $F_{20\%}$ – (bioavailability > 20%).
Skin permeability	-5.47	-5.56	-6.39	$\log Kp$ (cm/s).
Blood-brain barrier (BBB) penetration	0.039	0.107	0.023	The output value is the probability of being permeable (BBB+). Category 1: BBB+; Category 0: BBB
Acute oral toxicity	0.049	0.097	0.381	The output value is the probability of being highly toxic (Category 0: low–toxicity; Category 1: high–toxicity).
Toxicophore predictions				
Acute toxicity rule	0 alert	0 alert	0 alert	Predicted as toxic towards water sources based on 99 substructures
Non-biodegradable rule	0 alert	0 alert	0 alert	Predicted as non-biodegradable substances based on 19 substructures.
Skin sensitization rules	1 alert	1 alert	0 alert	Predicted as skin irritants based on 155 substructures.
Eye irritation	0.991	0.978	0.917	The output value is the probability of being irritants (Category 1: irritants; Category 0: non-irritants).
Environmental toxicity				
IGC ₅₀	-5.367	-5.796	-4.465	<i>Tetrahymena pyriformis</i> 50% growth inhibition concentration. The unit is $\log_{10}[(mg/L)/(MW \times 10^3)]$.
LC ₅₀ FM	-5.981	-6.466	-4.975	96 h fathead minnow 50% lethal concentration. The unit is $\log_{10}[(mg/L)/(MW \times 10^3)]$.
LC ₅₀ DM	-6.191	-6.547	-5.153	48 h <i>daphnia magna</i> 50% lethal concentration. The unit is $log_{10}[(mg/L)/(MW \times 10^3)]$.

Finally, the ecotoxicological characteristics were assessed using cheminformatics tools implemented in ADMETlab 2.0. The results indicated that they had irrelevant environmental toxicity, especially for three types of aquatic organisms, e.g., *tetrahymena pyriformis* (TP), fathead minnow (FM), and *daphnia magna* (DM). These toxic endpoints are a mandatory preliminary assessment of chemicals' acute toxicity on aquatic species, according to the European regulation on the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) [99]. As can be seen in Table 6, the lethal concentrations 50% (log unit) were determined to be from -4.5 to -6.6 mg/L, and NEW03 exhibited the lowest aquatic toxicity among the three compounds. The overall results obtained herein confirmed the appropriate organo-kinetic and biosafety level of the designed coumarins against human and environmental health, suggesting their suitability for industrial production as florescent chemosensors.

3.6. Theoretical Synthetic Routes Proposed for Designed Coumarins

Given the sensing characteristics' predictions provided by the QSPR and quantum chemical calculations, it is desirable to obtain the florescent probes with the most potential for the experimental evaluations. Even though chemical synthesis was out of the scope of the current study, a proposal of the theoretical synthetic routes may be very helpful to assess the developability of the proposed structures, as well as to serve as a guidance for further developing novel chemosensors. To this end, three coumarins, NEW02, NEW03, and NEW07, were selected for the synthesis predictions.

The synthetic probability and SA scores of NEW02, NEW03, and NEW07 were first determined using the chemoinformatic tools GASA, ADMETlab, and SwissADME [81,82,87]. The results are summarized in Table 7, suggesting that all three chemicals belong to the easy-to-synthesize class.

Table 7. Synthetic accessibility prediction for selected coumarins.

Cpd. ID	Synthetic Probability ¹	SwissADME SA Score ²	ADMET1ab 2.0 SA Score ³	Consensus Classification
NEW02	0.879	3.14	2.62	ES (easy-to-synthesize)
NEW03	0.900	3.17	3.01	ES (easy-to-synthesize)
NEW07	0.899	3.36	2.63	ES (easy-to-synthesize)

Synthetic accessibility: ¹ GASA synthetic probability ranges from 0 (hard-to-synthesize) to 1 (easy-to-synthesize); ² SwissADME SA scores range from 1 (very easy) to 10 (very difficult); ³ ADMETlab 2.0 SA score classification: SA score \geq 6, difficult-to-synthesize; SA score < 6, easy-to-synthesize.

Based on the synthetic accessibility predictions, we proposed in this study two theoretical synthetic routes: the first one for NEW02 and NEW03 and the other for NEW07. The first proposal, as depicted in Scheme 1, generates 7-amino-4-methyl coumarin (AMC) as a key intermediate, while the second one (Scheme 2) goes through another intermediate species, 7-amino-8-hydroxy-4-methylcoumarin (AHMC).



Scheme 1. Reagents and conditions: (a) Methyl chloroformate, $KHCO_3$, EtOAc, 4 h, r.t. (b) Ethyl 3-oxobutanoate, H_2SO_4 (70%), 18 h, r.t.. (c) H_2SO_4 (98%) and glacial acetic acid (1:1), reflux, 5 h. (d) EtOH, reflux, 10–12 h.

The theoretical synthetic routes proposed for NEW02 and NEW03 start with the preparation of methyl (3-hydroxyphenyl)carbamate (I) by acylation of the amino group in *m*-aminophenol with methyl chloroformate in the presence of a base (e.g., potassium bicarbonate). Subsequently, 7-carbethoxyamino-4-methylcoumarin (II) can be synthesized via Pechmann condensation with acetoacetic ester in sulfuric acid. Accordingly, the synthesis of AMC from II can be performed by heating II at reflux in H₂SO₄ (98%) and glacial acetic acid for around 5 h. After adjusting to a low temperature and to pH ~ 7, the precipitate is then dissolved in the ethanol solvent. The resulting ethanol solution can be used to obtain the target compounds (NEW02 or NEW03) by adding to other ethanol solutions containing 2-mercaptobenzaldehyde (IIIa) or 2,4-mercaptobenzaldehyde (IIIb), then the mixture must be refluxed for 10–12 h under stirring. Once finished, the end product is cooled to room temperature, and after that, it can be collected by filtration, dried, and purified by recrystallization from DMF and ethanol. Considering the previous reports [91],

pyrocatechol

IV

NEW02 and NEW03 can be produced in an overall yield greater than 50%. ÓН OH óн NEW07

AHMC

the intermediate AMC can be formed in good yield (~90%); thus, the recrystallized form of

Scheme 2. Reagents and conditions: (a) Ethyl 3-oxobutanoate, H₂SO₄ (70%), 24 h, 0 °C. (b) HNO₃/H₂SO₄, DCM, 0 °C, 1 h. (c) ZnCl₂, HCl, EtOAc, r.t. (d) NaBH₃CN/MeOH, AcOH, r.t., 1 h.

VI

The preparation of the remaining coumarin NEW07 based on the AHMC derivative is described by Scheme 2. The theoretical synthetic route is composed of four steps, starting with the formation of 8-hydroxy-4-methylcoumarin (IV) via a Pechmann condensation by reacting pyrocatechol and ethyl 3-oxobutanoate. Subsequently, nitration of IV with a mixture of nitric acid and sulfuric acid can lead to a separable mixture of both 7- and 8-nitrocoumarin derivatives. There are several methods proposed in the literature to reduce o-nitrohydroxy-4-methylcoumarins using different reducing agents [91,94]. In this study, 7amino-8-hydroxy-4-methylcoumarin (AHMC) was obtained by the reduction of 8-hydroxy-4-methyl-7-nitrocoumarin (V) using zinc chloride in hydrochloric acid and ethyl acetate to achieve a better yield. Finally, the target compound NEW07 can be obtained by reductive amination of AHMC with 2-hydroxybenzaldehyde using sodium cyanoborohydride in methanol and acetic acid (Scheme 2).

4. Conclusions

x

Heavy metal contamination in wastewater from industrial activities is escalating at alarming levels, and it has now become a major global problem. There has been a great necessity to effectively develop analytical methods for the identification and quantification of heavy metals in the aquatic medium. Florescent chemosensors have attracted major attention of researchers in this field due to their significant advantages related to visual simplicity, low cost, high sensitivity, good selectivity, and rapid response compared to the other methods. In assisting the rational design of potential chemosensors, computational approaches, e.g., the QSPR and DFT, have been widely developed and applied for a variety of heavy metals over the last few decades. In this study, a number of QSPR models were developed using the GA-MLR statistical and 10 machine learning techniques on a large database of ~200 florescent chemosensor structures in complexes with three heavy metals $(Cu^{2+}, Cd^{2+}, and Pb^{2+})$ previously reported. The modeling process clearly confirmed the ability of ML algorithms to improve the predictive performance of computational models. Six models based on GA-MLR, the AdaBoost regressor MLR, gradient boosting, and support vector regression, were selected and showed a good goodness-of-fit, robustness, and predictivity with the cross-validation accuracy ranging from >0.9 to 1.0, and those for external validation ranged from >0.8 to 0.95. It is worth noting that, so far, ML techniques have not been commonly used, then the predictive ability of the previously published studies appeared to be limited.

Subsequently, the developed computational models were applied for screening a virtual database of 129 coumarin-like structures, which were automatically generated based on seven potential florescent probes formerly reported in the literature. Interestingly, three novel simple compounds, namely NEW02, NEW03, and NEW07 were discovered with simultaneously high values of the stability constant and potentiometric sensitivity against three metal ions, Cu²⁺, Cd²⁺, and Pb²⁺. The QSPR predictions were then in good agreement with the quantum calculations using DFT/B3LYP approaches. All the HOMO–LUMO levels and other reactivity descriptors were in the same rankings as the QSPR approaches. To go further with the developability of the newly designed coumarin derivatives as chemosensors, their organo-kinetic and toxicological profiles were examined by computing

numerous physicochemical and several toxic endpoints against human and environmental health using several cheminformatics tools. The results once again confirmed their safety characteristics for both humans and other species in aquatic ecosystems. Finally, based on retrosynthesis analysis with computational tools, the theoretical synthetic proposal for the three coumarins designed was successfully provided. The compounds can be synthesized by an inexpensive and simple method with good yield. Based on the aforementioned findings, the newly designed coumarins could be considered for further development stages as potential candidates as chemosensors against three metal ions, Cu^{2+} , Cd^{2+} , and Pb^{2+} . The general workflow presented herein is promising, and we believe it can be applied to assist the development of novel chemosensors for screening multiple heavy metals' contamination, which is one of the major environmental problems facing our planet today.

Supplementary Materials: The following Supporting Information can be downloaded at: https:// www.mdpi.com/article/10.3390/toxics11070595/s1, Figure S1: Scatter plots of LMO validation approach; Figure S2: Y randomization test compared to the original QSPR model; Figures S3–S8: Correlation matrix of molecular descriptors selected in the six QSPR models; Table S1: Database of experimental potentiometric sensitivity (PS) of complexes between Cu^{2+} , Cd^{2+} , and Pb^{2+} ions and compounds; Tables S2–S4: Experimental stability constant (log β) of the complex between three metals (Cd^{2+} , Cu^{2+} , Pb^{2+}) towards compounds in the database; Tables S5–S10: Runtime parameters optimized for the ML algorithm predicting the log β and PS_{ML} of the complex between three metals (Cd^{2+} , Cu^{2+} , Pb^{2+}) and compounds in the database; Table S11: Molecular descriptors using in the GA-MLR models; Tables S12–S17: Performance comparison between the GA-MLR and ML models; Table S18: log β and PS_{ML} parameter predictions for virtual the chemical library based on the corresponding QSPR models; Table S19: SMILES code of the 8 new coumarin-like structures and their leverage value (h_i) computed by the 6 QSPR models; Tables S20–S26. Physicochemical and absorption, distribution, metabolism, and excretion (ADME) properties of three candidates based on the SwissADME tool.

Author Contributions: Conceptualization, Z.S., B.R., H.P.-T. and L.C.V.C.; methodology, B.R., G.M.C.-M., H.P.-T., C.-T.N. and L.C.V.C.; software, P.T.D.-T., T.-T.H., N.-V.T., L.-Q.B., H.T.P., T.T.G.C., H.T.B.M. and G.M.C.-M.; validation, Z.S., G.M.C.-M., T.-T.H., N.-V.T., L.-Q.B., H.T.P., T.T.G.C. and H.T.B.M.; formal analysis, B.R., H.P.-T. and L.C.V.C.; investigation, P.T.D.-T., T.-T.H., N.-V.T., L.-Q.B., H.T.P., T.T.G.C., H.T.B.M., C.-T.N. and G.M.C.-M.; resources, B.R., H.P.-T. and L.C.V.C.; data curation, P.T.D.-T., T.-T.H., N.-V.T. and L.-Q.B.; writing—original draft preparation, P.T.D.-T., T.-T.H. and L.-Q.B.; writing—review and editing, B.R., H.P.-T. and L.C.V.C.; visualization, T.-T.H., N.-V.T., L.-Q.B. and G.M.C.-M.; supervision, B.R., H.P.-T. and L.C.V.C.; project administration, P.T.D.-T. and L.C.V.C.; funding acquisition, L.C.V.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Vietnam Academy of Science and Technology under Grant Number CSCL36.01/22-23. This work was made possible in part by National Science Foundation (NSF) MRI Award No. 2019077.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data for model development and application are provided.

Acknowledgments: The authors thank Paola Gramatica for generously providing a free license for the QSARINS software. This work used resources of the Center for Computationally Assisted Science and Technology (CCAST) at North Dakota State University, which was made possible in part by National Science Foundation (NSF) MRI Award No. 2019077. Supercomputing support provided by the CCAST HPC System at NDSU is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of the data; in the writing of the manuscript; nor in the decision to publish the results.

References

- Romero-Estudillo, I.; Viveros-Ceballos, J.L.; Cazares-Carreño, O.; González-Morales, A.; de Jesús, B.F.; López-Castillo, M.; Razo-Hernández, R.S.; Castañeda-Corral, G.; Ordóñez, M. 000Synthesis of new α-aminophosphonates: Evaluation as antiinflammatory agents and QSAR studies. *Bioorg. Med. Chem.* 2019, 27, 2376–2386. [CrossRef] [PubMed]
- Lin, L.; Yang, H.; Xu, X. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. *Front. Environ. Sci.* 2022, 10, 880246. [CrossRef]
- 3. Noor, R.; Maqsood, A.; Baig, A.; Pande, C.B.; Zahra, S.M.; Saad, A.; Anwar, M.; Singh, S.K. A comprehensive review on water pollution, South Asia Region: Pakistan. *Urban Clim.* **2023**, *48*, 101413. [CrossRef]
- Md Anawar, H.; Chowdhury, R. Remediation of Polluted River Water by Biological, Chemical, Ecological and Engineering Processes. Sustainability 2020, 12, 7017. [CrossRef]
- 5. Vhahangwele, M.; Khathutshelo, L.M. Environmental Contamination by Heavy Metals. In *Heavy Metals*; Hosam El-Din, M.S., Refaat, F.A., Eds.; IntechOpen: Rijeka, Croatia, 2018; pp. 115–133.
- 6. Järup, L. Hazards of heavy metal contamination. Br. Med. Bull. 2003, 68, 167–182. [CrossRef]
- 7. Tchounwou, P.B.; Yedjou, C.G.; Patlolla, A.K.; Sutton, D.J. Heavy Metal Toxicity and the Environment. In *Molecular, Clinical and Environmental Toxicology: Volume 3: Environmental Toxicology;* Luch, A., Ed.; Springer: Basel, Switzerland, 2012; pp. 133–164.
- 8. Rafati-Rahimzadeh, M.; Rafati-Rahimzadeh, M.; Kazemi, S.; Moghadamnia, A. Cadmium toxicity and treatment: An update. *Caspian J. Intern. Med.* **2017**, *8*, 135–145. [CrossRef]
- 9. Ebrahimi, M.; Khalili, N.; Razi, S.; Keshavarz-Fathi, M.; Khalili, N.; Rezaei, N. Effects of lead and cadmium on the immune system and cancer progression. *J. Environ. Health Sci. Eng.* **2020**, *18*, 335–343. [CrossRef]
- 10. Flemming, C.A.; Trevors, J.T. Copper toxicity and chemistry in the environment: A review. *Water Air Soil Pollut.* **1989**, *44*, 143–158. [CrossRef]
- 11. Cao, R.; Zhang, Y.; Ju, Y.; Wang, W.; Xi, C.; Liu, W.; Liu, K. Exacerbation of copper pollution toxicity from ocean acidification: A comparative analysis of two bivalve species with distinct sensitivities. *Environ. Pollut.* **2022**, 293, 118525. [CrossRef]
- Krämer, J.; Kang, R.; Grimm, L.M.; De Cola, L.; Picchetti, P.; Biedermann, F. Molecular Probes, Chemosensors, and Nanosensors for Optical Detection of Biorelevant Molecules and Ions in Aqueous Media and Biofluids. *Chem. Rev.* 2022, 122, 3459–3636. [CrossRef]
- 13. Punia, P.; Bharti, M.K.; Dhar, R.; Thakur, P.; Thakur, A. Recent Advances in Detection and Removal of Heavy Metals from Contaminated Water. *ChemBioEng Rev.* 2022, *9*, 351–369. [CrossRef]
- 14. Wu, D.; Sedgwick, A.C.; Gunnlaugsson, T.; Akkaya, E.U.; Yoon, J.; James, T.D. Fluorescent chemosensors: The past, present and future. *Chem. Soc. Rev.* 2017, *46*, 7105–7123. [CrossRef]
- 15. Dongare, P.R.; Gore, A.H. Recent Advances in Colorimetric and Fluorescent Chemosensors for Ionic Species: Design, Principle and Optical Signalling Mechanism. *ChemistrySelect* **2021**, *6*, 5657–5669. [CrossRef]
- 16. Vladimirova, N.; Polukeev, V.; Ashina, J.; Babain, V.; Legin, A.; Kirsanov, D. Prediction of Carbonate Selectivity of PVC-Plasticized Sensor Membranes with Newly Synthesized Ionophores through QSPR Modeling. *Chemosensors* **2022**, *10*, 43. [CrossRef]
- 17. Lieberzeit, P.A.; Dickert, F.L. Chemosensors in environmental monitoring: Challenges in ruggedness and selectivity. *Anal. Bioanal. Chem.* **2009**, *393*, 467–472. [CrossRef]
- Madatov, U.; Rakhimov, S.; Shahidova, D.; Smanova, Z.; Lal, B.; Berdimurodov, E. A new, green, highly effective procedure for manganese determination using alizarin-3-methylamino-N,N-diacetic acid immobilised on a polymer matrix. *Int. J. Environ. Anal. Chem.* 2022, 1–22. [CrossRef]
- 19. Smanova, Z.A.; Usmanova, X.U.; Madusmanova, N.K.; Bobojanov, B.B.; Maxmudova, G.O. Immobilized Oxyazo Compounds as Analytical Reagents for the Sorption-Luminescent Determination of Certain Metals. *Turk. Online J. Qual. Ing.* **2021**, *12*, 6113–6119.
- Fjodorova, N.; Novic, M.; Gajewicz, A.; Rasulev, B. The way to cover prediction for cytotoxicity for all existing nano-sized metal oxides by using neural network method. *Nanotoxicology* 2017, 11, 475–483. [CrossRef]
- Toropov, A.A.; Toropova, A.P.; Rasulev, B.F.; Benfenati, E.; Gini, G.; Leszczynska, D.; Leszczynski, J. Coral: QSPR modeling of rate constants of reactions between organic aromatic pollutants and hydroxyl radical. J. Comput. Chem. 2012, 33, 1902–1906. [CrossRef]
- Sizochenko, N.; Syzochenko, M.; Fjodorova, N.; Rasulev, B.; Leszczynski, J. Evaluating genotoxicity of metal oxide nanoparticles: Application of advanced supervised and unsupervised machine learning techniques. *Ecotoxicol. Environ. Saf.* 2019, 185, 109733. [CrossRef]
- 23. Rasulev, B.; Kušić, H.; Leszczynska, D.; Leszczynski, J.; Koprivanac, N. QSAR modeling of acute toxicity on mammals caused by aromatic compounds: The case study using oral LD50 for rats. *J. Environ. Monit.* **2010**, *12*, 1037. [CrossRef] [PubMed]
- Turabekova, M.A.; Rasulev, B.F.; Dzhakhangirov, F.N.; Leszczynska, D.; Leszczynski, J. Aconitum and Delphinium alkaloids of curare-like activity. QSAR analysis and molecular docking of alkaloids into AChBP. *Eur. J. Med. Chem.* 2010, 45, 3885–3894. [CrossRef] [PubMed]
- 25. Gooch, A.; Sizochenko, N.; Rasulev, B.; Gorb, L.; Leszczynski, J. In vivo toxicity of nitroaromatics: A comprehensive quantitative structure–activity relationship study. *Environ. Toxicol. Chem.* **2017**, *36*, 2227–2233. [CrossRef] [PubMed]
- Ahmed, L.; Rasulev, B.; Kar, S.; Krupa, P.; Mozolewska, M.A.; Leszczynski, J. Inhibitors or toxins? Large library target-specific screening of fullerene-based nanoparticles for drug design purpose. *Nanoscale* 2017, 9, 10263–10276. [CrossRef]

- Ukić, Š.; Sigurnjak, M.; Cvetnić, M.; Markić, M.; Stankov, M.N.; Rogošić, M.; Rasulev, B.; Lončarić Božić, A.; Kušić, H.; Bolanča, T. Toxicity of pharmaceuticals in binary mixtures: Assessment by additive and non-additive toxicity models. *Ecotoxicol. Environ. Saf.* 2019, 185, 109696. [CrossRef]
- Turabekova, M.A.; Rasulev, B.F.; Dzhakhangirov, F.N.; Salikhov, S.I. Aconitum and Delphinium alkaloids: "Drug-likeness" descriptors related to toxic mode of action. *Environ. Toxicol. Pharmacol.* 2008, 25, 310–320. [CrossRef]
- 29. Soloviev, V.; Varnek, A.; Babain, V.; Polukeev, V.; Ashina, J.; Legin, E.; Legin, A.; Kirsanov, D. QSPR modeling of potentiometric sensitivity towards heavy metal ions for polymeric membrane sensors. *Sens. Actuators B Chem.* **2019**, *301*, 126941. [CrossRef]
- Vladimirova, N.; Puchkova, E.; Dar'in, D.; Turanov, A.; Babain, V.; Kirsanov, D. Predicting the Potentiometric Sensitivity of Membrane Sensors Based on Modified Diphenylphosphoryl Acetamide Ionophores with QSPR Modeling. *Membranes* 2022, 12, 953. [CrossRef]
- 31. Kiani-Anbouhi, R.; Ganjali, M.R.; Norouzi, P. Prediction of the complexation stabilities of La³⁺ ion with ionophores applied in lanthanoid sensors. *J. Incl. Phenom. Macrocycl. Chem.* **2014**, *78*, 325–336. [CrossRef]
- 32. Solov'ev, V.; Kireeva, N.; Ovchinnikova, S.; Tsivadze, A. The complexation of metal ions with various organic ligands in water: Prediction of stability constants by QSPR ensemble modelling. *J. Incl. Phenom. Macrocycl. Chem.* **2015**, *83*, 89–101. [CrossRef]
- Martynko, E.; Solov'ev, V.; Varnek, A.; Legin, A.; Kirsanov, D. QSPR Modeling of Potentiometric Mg²⁺/Ca²⁺ Selectivity for PVC-plasticized Sensor Membranes. *Electroanalysis* 2020, *32*, 792–798. [CrossRef]
- 34. Kiani-Anbouhi, R.; Ganjali, M.R.; Norouzi, P. Application of QSPR for prediction of the complexation stabilities of Sm(III) with ionophores applied in lanthanoid sensors. *J. Incl. Phenom. Macrocycl. Chem.* **2015**, *81*, 441–450. [CrossRef]
- Kanahashi, K.; Urushihara, M.; Yamaguchi, K. Machine learning-based analysis of overall stability constants of metal–ligand complexes. Sci. Rep. 2022, 12, 11159. [CrossRef]
- Solov'ev, V.; Sukhno, I.; Buzko, V.; Polushin, A.; Marcou, G.; Tsivadze, A.; Varnek, A. Stability constants of complexes of Zn²⁺, Cd²⁺, and Hg²⁺ with organic ligands: QSPR consensus modeling and design of new metal binders. *J. Incl. Phenom. Macrocycl. Chem.* 2012, 72, 309–321. [CrossRef]
- 37. Yamasaki, K.; Yasuda, M. Stability of Zinc and Cadmium Complexes with 2,2'-Bipyridine and 1,10-Phenanthroline. *J. Am. Chem. Soc.* **1956**, *78*, 1324. [CrossRef]
- Golcu, A.; Tumer, M.; Demirelli, H.; Wheatley, R.A. Cd(II) and Cu(II) complexes of polydentate Schiff base ligands: Synthesis, characterization, properties and biological activity. *Inorg. Chim. Acta* 2005, 358, 1785–1797. [CrossRef]
- Sóvágó, I.; Várnagy, K. Cadmium(II) Complexes of Amino Acids and Peptides. In *Cadmium: From Toxicity to Essentiality*; Sigel, A., Sigel, H., Sigel, R.K.O., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 275–302.
- Quang, N.M.; Mau, T.X.; Ai Nhung, N.T.; Minh An, T.N.; Van Tat, P. Novel QSPR modeling of stability constants of metalthiosemicarbazone complexes by hybrid multivariate technique: GA-MLR, GA-SVR and GA-ANN. J. Mol. Struct. 2019, 1195, 95–109. [CrossRef]
- Hancock, R.D.; de Sousa, A.S.; Walton, G.B.; Reibenspies, J.H. Metal-Ion Selectivity Produced by C-Alkyl Substituents on the Bridges of Chelating Ligands: The Importance of Short H–H Nonbonded van der Waals Contacts in Controlling Metal-Ion Selectivity. A Thermodynamic, Molecular Mechanics, and Crystallographic Study. *Inorg. Chem.* 2007, 46, 4749–4757. [CrossRef]
- 42. Kotek, J.; Kubíček, V.; Hermann, P.; Lukeš, I. Synthesis and Characterization of Ligands and their Gadolinium(III) Complexes. In *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*; John Wiley & Sons, Ltd.: West Sussex, UK, 2013; pp. 83–155.
- Sigel, H.; Da Costa, C.P.; Song, B.; Carloni, P.; Gregáň, F. Stability and Structure of Metal Ion Complexes Formed in Solution with Acetyl Phosphate and Acetonylphosphonate: Quantification of Isomeric Equilibria. J. Am. Chem. Soc. 1999, 121, 6248–6257. [CrossRef]
- Fernández-Botello, A.; Griesser, R.; Holý, A.; Moreno, V.; Sigel, H. Acid–Base and Metal-Ion-Binding Properties of 9-[2-(2-Phosphonoethoxy)ethyl]adenine (PEEA), a Relative of the Antiviral Nucleotide Analogue 9-[2-(Phosphonomethoxy)ethyl]adenine (PMEA). An Exercise on the Quantification of Isomeric Complex Equilibria in Solution. *Inorg. Chem.* 2005, 44, 5104–5117. [CrossRef]
- 45. Hardy, J.G. Metallosupramolecular grid complexes: Towards nanostructured materials with high-tech applications. *Chem. Soc. Rev.* **2013**, *42*, 7881–7899. [CrossRef] [PubMed]
- Kaczorowska, M.A.; Bożejewicz, D.; Witt, K.; Urbaniak, W. A new application of 2-benzoylpyridine–efficient removal of silver ions from acidic aqueous solutions via adsorption process on polymeric material and classic solvent extraction. *Chem. Process Eng.* 2022, 43, 369–382.
- Jung, H.S.; Kwon, P.S.; Lee, J.W.; Kim, J.I.; Hong, C.S.; Kim, J.W.; Yan, S.; Lee, J.Y.; Lee, J.H.; Joo, T.; et al. Coumarin-Derived Cu²⁺-Selective Fluorescence Sensor: Synthesis, Mechanisms, and Applications in Living Cells. *J. Am. Chem. Soc.* 2009, 131, 2008–2012. [CrossRef] [PubMed]
- Agarwal, R.P.; Perrin, D.D. Stability constants of complexes of copper(II) ions with some histidine peptides. *Dalton Trans.* 1975, 3, 268–272. [CrossRef]
- 49. Le Fur, M.; Beyler, M.; Le Poul, N.; Lima, L.M.P.; Le Mest, Y.; Delgado, R.; Platas-Iglesias, C.; Patinec, V.; Tripier, R. Improving the stability and inertness of Cu(ii) and Cu(i) complexes with methylthiazolyl ligands by tuning the macrocyclic structure. *Dalton Trans.* **2016**, *45*, 7406–7420. [CrossRef]
- Nagaj, J.; Stokowa-Sołtys, K.; Kurowska, E.; Frączyk, T.; Jeżowska-Bojczuk, M.; Bal, W. Revised Coordination Model and Stability Constants of Cu(II) Complexes of Tris Buffer. *Inorg. Chem.* 2013, 52, 13927–13933. [CrossRef] [PubMed]

- 51. Royzen, M.; Dai, Z.; Canary, J.W. Ratiometric Displacement Approach to Cu(II) Sensing by Fluorescence. *J. Am. Chem. Soc.* 2005, 127, 1612–1613. [CrossRef] [PubMed]
- Xie, J.; Ménand, M.; Maisonneuve, S.; Métivier, R. Synthesis of Bispyrenyl Sugar-Aza-Crown Ethers as New Fluorescent Molecular Sensors for Cu(II). J. Org. Chem. 2007, 72, 5980–5985. [CrossRef]
- Tanaka, M.; Tabata, M. Stability Constants of Metal(II) Complexes with Amines and Aminocarboxylates with Special Reference to Chelation. Bull. Chem. Soc. Jpn. 2009, 82, 1258–1265. [CrossRef]
- 54. Türkel, N. Equilibrium Study of the Mixed Complexes of Copper(II) with Adenine and Amino Acids in Aqueous Solution. *J. Solut. Chem.* **2015**, *44*, 1267–1280. [CrossRef]
- 55. Agrawal, B.R.; Magare, B.K.; Farooqui, M.N.; Janrao, D.M.; Ubale, M.B. Stability Constants of Cu (II), Ni (II) and Mn (II) Metal Complexes with Cetrizine and Benzoic acid. *Int. J. Chem. Sci.* 2009, 7, 2169–2172.
- Pashkina, D.A.; Gusev, V.Y.; Radushev, A.V. Complexation of copper(II) with 2',2'-dialkyl-para-tert-butylbenzohydrazides. *Russ. J. Inorg. Chem.* 2014, 59, 394–398. [CrossRef]
- 57. Hay, R.W.; Govan, N.; Perotti, A.; Carugo, O. Copper(II), nickel(II) and palladium(II) complexes of the diamide ligand N,N'bis(2-carbamoylethyl)ethylenediamine (H₂L) and the crystal structure of the carbonyl-oxygen-bonded copper(II) complex [Cu(H₂L)](ClO₄)₂. *Transit. Met. Chem.* **1997**, *22*, 389–394. [CrossRef]
- Polat, F.; Atabey, H.; Sari, H.; Cukurovali, A. Potentiometric study of equilibrium constants of a novel triazine-thione derivative and its stability constants with Hg²⁺, Cu²⁺, Ni²⁺, Pb²⁺, and Zn²⁺ metal ions in ethanol and water mixed. *Turk. J. Chem.* 2013, 37, 439–448. [CrossRef]
- Kiani-anboui, R.; Ghasemi, M.H. A QSPR Study for the Prediction of the Selectivity of Pb(II) Sensors by Stability Constants of Ion-Ionophore Complexes. *Anal. Bioanal. Electrochem.* 2022, 14, 598–609.
- 60. Jagvir, S.; Abhay Nanda, S.; Netrapal, S.; Anuradha, S. Stability Constants of Metal Complexes in Solution. In *Stability and Applications of Coordination Compounds*; Abhay Nanda, S., Ed.; IntechOpen: Rijeka, Croatia, 2019; pp. 1–18.
- 61. Todeschini, R.; Consonni, V. Molecular Descriptors for Chemoinformatics; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009; Volume 11, p. 967.
- 62. Perez-Castillo, Y.; Sánchez-Rodríguez, A.; Tejera, E.; Cruz-Monteagudo, M.; Borges, F.; Cordeiro, M.N.D.S.; Le-Thi-Thu, H.; Pham-The, H. A desirability-based multi objective approach for the virtual screening discovery of broad-spectrum anti-gastric cancer agents. *PLoS ONE* **2018**, *13*, e0192176. [CrossRef]
- 63. Pham-The, H.; Casañola-Martin, G.; Diéguez-Santana, K.; Nguyen-Hai, N.; Ngoc, N.T.; Vu-Duc, L.; Le-Thi-Thu, H. Quantitative structure-activity relationship analysis and virtual screening studies for identifying HDAC2 inhibitors from known HDAC bioactive chemical libraries. *SAR QSAR Environ. Res.* **2017**, *28*, 199–220. [CrossRef]
- 64. Daghighi, A.; Casanola-Martin, G.M.; Timmerman, T.; Milenković, D.; Lučić, B.; Rasulev, B. In Silico Prediction of the Toxicity of Nitroaromatic Compounds: Application of Ensemble Learning QSAR Approach. *Toxics* **2022**, *10*, 746. [CrossRef]
- 65. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem.* **2013**, *34*, 2121–2132. [CrossRef]
- 66. Pham-The, H.; Le-Thi-Thu, H. Integrating structure and ligand-based approaches for modelling the Histone deacetylase inhibition activity of hydroxamic acid derivatives. *Asian J. Pharm. Clin. Res.* **2018**, *11*, 198–206. [CrossRef]
- 67. Golbraikh, A.; Tropsha, A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *Mol. Divers.* 2000, *5*, 231–243. [CrossRef]
- Nimon, K.F.; Oswald, F.L. Understanding the Results of Multiple Linear Regression: Beyond Standardized Regression Coefficients. Organ. Res. Methods 2013, 16, 650–674. [CrossRef]
- 69. Jerome, H.F. Multivariate Adaptive Regression Splines. Ann. Statist. 1991, 19, 1–67. [CrossRef]
- 70. Bondarchuk, S.V. Friction sensitivity of nitramine energetic materials: A prediction based on genetic function approximation. *FirePhysChem* **2022**, *2*, 272–278. [CrossRef]
- Bondarchuk, S.V. Prediction of aquatic toxicity of energetic materials using genetic function approximation. *FirePhysChem* 2023, 3, 23–28. [CrossRef]
- Gramatica, P. On the Development and Validation of QSAR Models. In *Computational Toxicology: Volume II*; Reisfeld, B., Mayeno, A.N., Eds.; Humana Press: Totowa, NJ, USA, 2013; pp. 499–526.
- 73. Gramatica, P. Principles of QSAR models validation: Internal and external. QSAR Comb. Sci. 2007, 26, 694–701. [CrossRef]
- 74. Netzeva, T.I.; Worth, A.P.; Aldenberg, T.; Benigni, R.; Cronin, M.T.D.; Gramatica, P.; Jaworska, J.S.; Kahn, S.; Klopman, G.; Marchant, C.A.; et al. Current Status of Methods for Defining the Applicability Domain of (Quantitative) Structure-Activity Relationships: The Report and Recommendations of ECVAM Workshop 521,2. *Altern. Lab. Anim.* 2005, *33*, 155–173. [CrossRef]
- 75. Cao, D.; Liu, Z.; Verwilst, P.; Koo, S.; Jangjili, P.; Kim, J.S.; Lin, W. Coumarin-Based Small-Molecule Fluorescent Chemosensors. *Chem. Rev.* 2019, 119, 10403–10519. [CrossRef]
- 76. Gill, P.M.W.; Johnson, B.G.; Pople, J.A.; Frisch, M.J. The performance of the Becke—Lee—Yang—Parr (B—LYP) density functional theory with various basis sets. *Chem. Phys. Lett.* **1992**, 197, 499–505. [CrossRef]
- 77. Guest, M.F.; Bush, I.J.; Van Dam, H.J.J.; Sherwood, P.; Thomas, J.M.H.; Van Lenthe, J.H.; Havenith, R.W.A.; Kendrick, J. The GAMESS-UK electronic structure package: Algorithms, developments and applications. *Mol. Phys.* 2005, 103, 719–747. [CrossRef]
- 78. Parr, R.G.; Donnelly, R.A.; Levy, M.; Palke, W.E. Electronegativity: The density functional viewpoint. J. Chem. Phys. 1978, 68, 3801–3807. [CrossRef]

- 79. Makov, G. Chemical Hardness in Density Functional Theory. J. Phys. Chem. 1995, 99, 9337–9339. [CrossRef]
- 80. Parr, R.G.; Szentpály, L.v.; Liu, S. Electrophilicity Index. J. Am. Chem. Soc. 1999, 121, 1922–1924. [CrossRef]
- 81. Daina, A.; Michielin, O.; Zoete, V. SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 2017, *7*, 42717. [CrossRef] [PubMed]
- 82. Xiong, G.; Wu, Z.; Yi, J.; Fu, L.; Yang, Z.; Hsieh, C.; Yin, M.; Zeng, X.; Wu, C.; Lu, A.; et al. ADMETlab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* **2021**, *49*, W5–W14. [CrossRef]
- 83. Castillo-Garit, J.A.; Casanola-Martin, G.M.; Le-Thi-Thu, H.; Pham-The, H.; Barigye, S.J. A Simple Method to Predict Blood-Brain Barrier Permeability of Drug- Like Compounds Using Classification Trees. *Med. Chem.* **2017**, *13*, 664–669. [CrossRef]
- Pham-The, H.; Garrigues, T.; Bermejo, M.; González-Álvarez, I.; Monteagudo, M.C.; Cabrera-Pérez, M. Provisional classification and in silico study of biopharmaceutical system based on caco-2 cell permeability and dose number. *Mol. Pharm.* 2013, 10, 2445–2461. [CrossRef]
- 85. Dieguez-Santana, K.; Pham-The, H.; Villegas-Aguilar, P.J.; Le-Thi-Thu, H.; Castillo-Garit, J.A.; Casañola-Martin, G.M. Prediction of acute toxicity of phenol derivatives using multiple linear regression approach for Tetrahymena pyriformis contaminant identification in a median-size database. *Chemosphere* **2016**, *165*, 434–441. [CrossRef]
- 86. Pham-The, H.; González-Alvarez, I.; Bermejo, M.; Garrigues, T.; Le-Thi-Thu, H.; Cabrera-Pérez, M.A. The Use of Rule-Based and QSPR Approaches in ADME Profiling: A Case Study on Caco-2 Permeability. *Mol. Inf.* **2013**, *32*, 459–479. [CrossRef]
- Yu, J.; Wang, J.; Zhao, H.; Gao, J.; Kang, Y.; Cao, D.; Wang, Z.; Hou, T. Organic Compound Synthetic Accessibility Prediction Based on the Graph Attention Mechanism. J. Chem. Inf. Model. 2022, 62, 2973–2986. [CrossRef]
- 88. Ertl, P.; Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **2009**, *1*, 8. [CrossRef]
- 89. Austin, P.C.; Steyerberg, E.W. The number of subjects per variable required in linear regression analyses. *J. Clin. Epidemiol.* **2015**, 68, 627–636. [CrossRef] [PubMed]
- Pham-The, H.; Nam, N.-H.; Nga, D.-V.; Hai, T.D.; Dieguez-Santana, K.; Marrero-Ponce, Y.; Castillo-Garit, A.J.; Casanola-Martin, M.G.; Le-Thi-Thu, H. Learning from Multiple Classifier Systems: Perspectives for Improving Decision Making of QSAR Models in Medicinal Chemistry. *Curr. Top. Med. Chem.* 2017, *17*, 3269–3288. [CrossRef] [PubMed]
- 91. Sun, X.-y.; Liu, T.; Sun, J.; Wang, X.-j. Synthesis and application of coumarin fluorescence probes. *RSC Adv.* **2020**, *10*, 10826–10847. [CrossRef] [PubMed]
- Sabeel, M.B.; Puthiyavalappil, R.; Vipin, M.; Vishnunarayanan Namboothiri Vadakkedathu, P.; Anandaram, S. Recent Advancements in Schiff Base Fluorescence Chemosensors for the Detection of Heavy Metal Ions. In Schiff Base in Organic, Inorganic and Physical Chemistry; Takashiro, A., Ed.; IntechOpen: Rijeka, Croatia, 2023; pp. 1–22.
- 93. Fan, L. Synthesis of Two Coumarin-Derived Schiff Bases and Investigation of theirs Selectivity for Zn²⁺. J. Fluoresc. 2017, 27, 1331–1337. [CrossRef]
- Chang, H.-Q.; Zhao, X.-L.; Wu, W.-N.; Jia, L.; Wang, Y. A highly sensitive on-off fluorescent chemosensor for Cu²⁺ based on coumarin. *J. Lumin.* 2017, 182, 268–273. [CrossRef]
- 95. Amalia, S.; Lucia, P. Molecular Descriptors and Properties of Organic Molecules. In *Symmetry (Group Theory) and Mathematical Treatment in Chemistry*; Takashiro, A., Ed.; IntechOpen: Rijeka, Croatia, 2018; pp. 161–176.
- 96. Karaca, H.; Kazancı, S. The metal sensing applications of chalcones: The synthesis, characterization and theoretical calculations. *J. Mol. Struct.* **2022**, *1248*, 131454. [CrossRef]
- Askerov, R.K.; El Bakri, Y.; Osmanov, V.K.; Ahmad, S.; Saravanan, K.; Borisova, G.N.; Nazarov, R.H.o.; Baranov, E.V.; Fukin, G.K.; Fukina, D.G.; et al. Complexes of 1-(2-R(F, CH₃, Cl)-phenyl)-1,4-dihydro-5H-tetrazole-5-thiones with cadmium chloride: Synthesis, molecular, crystal structures and computational investigation approach. *J. Inorg. Biochem.* 2022, 231, 111791. [CrossRef]
- 98. Cabrera-Perez, M.A.; Bermejo, M.; Alvarez, I.G.; Alvarez, M.G.; Garrigues, T.M. QSPR in Oral Bioavailability: Specificity or Integrality? *Mini-Rev. Med. Chem.* 2012, 12, 534–550. [CrossRef]
- 99. Tinkov, O.; Polishchuk, P.; Matveieva, M.; Grigorev, V.; Grigoreva, L.; Porozov, Y. The Influence of Structural Patterns on Acute Aquatic Toxicity of Organic Compounds. *Mol. Inf.* 2021, *40*, 2000209. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.