




## Article

# Towards Rational Biosurfactant Design—Predicting Solubilization in Rhamnolipid Solutions

Ilona E. Kłosowska-Chomiczewska <sup>1,\*</sup>, Adrianna Kotewicz-Siudowska <sup>1</sup>, Wojciech Artichowicz <sup>2</sup>, Adam Macierzanka <sup>1</sup>, Agnieszka Głowacz-Różyńska <sup>1</sup>, Patrycja Szumała <sup>1</sup>, Krystyna Mędrzycka <sup>1</sup>, Elżbieta Hallmann <sup>1</sup>, Elena Karpenko <sup>3</sup> and Christian Jungnickel <sup>1</sup>

<sup>1</sup> Department of Colloid and Lipid Science, Faculty of Chemistry, Gdańsk University of Technology, Narutowicza St. 11/12, 80-233 Gdańsk, Poland; adriannakotewicz@gmail.com (A.K.-S.); adamacie@pg.edu.pl (A.M.); agnglowa@pg.gda.pl (A.G.-R.); patszuma@pg.edu.pl (P.S.); kbm@pg.edu.pl (K.M.); hallmann.elzbieta@gmail.com (E.H.); cjungnickel@googlemail.com (C.J.)

<sup>2</sup> Department of Hydraulic Engineering, Faculty of Civil and Environmental Engineering, Gdańsk University of Technology, Narutowicza St. 11/12, 80-233 Gdańsk, Poland; wojart@pg.edu.pl

<sup>3</sup> Department of Physical Chemistry of Fossil Fuels InPOCC, National Academy of Sciences of Ukraine, 3a Naukova St., 79053 Lviv, Ukraine; e.v.karpenko@gmail.com

\* Correspondence: ilocho@pg.edu.pl; Tel.: +48-58-347-1151



**Citation:** Kłosowska-Chomiczewska, I.E.; Kotewicz-Siudowska, A.; Artichowicz, W.; Macierzanka, A.; Głowacz-Różyńska, A.; Szumała, P.; Mędrzycka, K.; Hallmann, E.; Karpenko, E.; Jungnickel, C. Towards Rational Biosurfactant Design—Predicting Solubilization in Rhamnolipid Solutions. *Molecules* **2021**, *26*, 534. <https://doi.org/10.3390/molecules26030534>

Academic Editor:

Lukasz Chrzanowski

Received: 19 November 2020

Accepted: 15 January 2021

Published: 20 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/>).

**Abstract:** The efficiency of micellar solubilization is dictated inter alia by the properties of the solubilize, the type of surfactant, and environmental conditions of the process. We, therefore, hypothesized that using the descriptors of the aforementioned features we can predict the solubilization efficiency, expressed as molar solubilization ratio (MSR). In other words, we aimed at creating a model to find the optimal surfactant and environmental conditions in order to solubilize the substance of interest (oil, drug, etc.). We focused specifically on the solubilization in biosurfactant solutions. We collected data from literature covering the last 38 years and supplemented them with our experimental data for different biosurfactant preparations. Evolutionary algorithm (EA) and kernel support vector machines (KSVM) were used to create predictive relationships. The descriptors of biosurfactant ( $\log P_{BS}$ , measure of purity), solubilize ( $\log P_{sol}$ , molecular volume), and descriptors of conditions of the measurement (T and pH) were used for modelling. We have shown that the MSR can be successfully predicted using EAs, with a mean  $R^2_{val}$  of  $0.773 \pm 0.052$ . The parameters influencing the solubilization efficiency were ranked upon their significance. This represents the first attempt in literature to predict the MSR with the MSR calculator delivered as a result of our research.

**Keywords:** rhamnolipid; biosurfactant; micellar solubilization; efficiency; MSR; QSAR; prediction

## 1. Introduction

Next to wetting, mobilization, and emulsification, solubilization is a common mechanism supporting a variety of domestic and industrial processes, starting from everyday hygiene, washing dishes and doing laundry, including environmental protection and drug delivery systems. For example, in environmental protection, solubilization is an inherent stage of water and ground remediation [1–3], which results in a significant increase of pollutants bioavailability for compatible microorganisms [4], therefore supporting self-purification or engineered purification of contaminated sites. In regard to drug delivery systems, solubilization serves not only for the preparation of normally water insoluble drug formulations, but also for their targeted delivery, e.g., in anti-cancer therapies [5,6]. Moreover, solubilization forms an important aspect of the efficient functioning of our digestive system, where human surfactants, i.e., bile salts solubilize consumed fat in order to increase the interfacial surface for enzymatic action of lipase/colipase complex [7].

For any given solubilization application it is crucial to reduce the environmental and human impact, and at the same time provide high efficiency. The human impact of

solubilization is especially pronounced by the surfactants' tendency to irritate human skin in domestic applications, whereas the environmental impact is mostly characterized by their resistance to biodegradation [8,9]. Therefore, an alternative to synthetic surfactants is commonly found among natural surfactants, i.e., biosurfactants, with rhamnolipids (RLs) being the most widely examined and commercially available group [10,11].

RLs are glycolipids produced by variety of microorganisms, often pathogenic, to support foundations of microbial life, i.e., food absorption or biofilm development [12,13]. For industrial purposes however, special emphasis is paid to biosurfactant production by non-pathogenic strains [14–17]. It allows not only to meet health and safety requirements in bioprocessing, but also to satisfy high requirements for raw materials for pharmaceuticals and cosmetics production. RLs are considered being active in a wide range of pH [18,19] and salinity [19], low- or non-toxic [20], easily biodegradable [20,21], and mild to the skin [22,23]. Except positive effects of RLs, there were reported cases where the biosurfactant either had no impact or even influenced the process negatively. The latter was rather matrix-dependent, i.e., resulted from the presence of co-contaminants next to RLs [24]. Such cases concerned e.g., preferential use of RL as a carbon source by bacteria and resulting lack of stimulation of hydrocarbon degradation [25] or increase of diesel oil phytotoxicity [26]. Notwithstanding, these do not undermine the general overtone of literature, bringing RLs to the position of green surfactants. Not without significance are the facts that CMC of RLs is low compared to synthetic surfactants (Table S1), and they were reported to solubilize hydrocarbons at concentrations even below the CMC [27]. These altogether make RL biosurfactants suitable for many domestic and industrial applications, implying they can be applied at reduced quantities, compared to synthetic surfactants. In spite of vast number of research on efficient methods for RLs production [28–30], the overall cost of the synthesis is high, and therefore limits the potential to widely replace surfactants with petrochemical origin. In the scientific world, multidimensional efforts are made to overcome this obstacle [31].

The efficiency of solubilization is most commonly presented in the form of weight or molar solubilization ratio (WSR and MSR, respectively) and is expressed as the ratio of the amount of solubilize to the amount of surfactant used, as shown in Equation (1):

$$\text{MSR} = \frac{C - C_{\text{CMC}}}{C_{\text{surf}} - \text{CMC}} \quad (1)$$

where  $C$  is the concentration of solubilized compound in a micellar solution,  $C_{\text{CMC}}$  is the solubility of solubilized compound in a micellar solution at the CMC,  $C_{\text{surf}}$  is the concentration of the surfactant, and CMC is the critical micellar concentration of the surfactant. For instance, the MSR for *n*-dodecane varied from 0.27 [32] to even 2.91 [27], and the MSR for *n*-hexadecane was 5.2 [33].

The MSR may vary upon different factors connected with the rhamnolipid itself, the solubilize, but also upon the experimental/industrial conditions. The MSR is crucial for the application of biosurfactants for several purposes in the industry. Being able to predict the MSR, one could more easily find an efficient surfactant for target application and limit the number of experiments needed for the research. In case of rhamnolipids, this could also be a step forward to design biosynthesis of these biomolecules, as the process usually results in obtaining a mixture of different congeners. The congeners differ in basic physicochemical properties, e.g., CMC or the ability to stabilize emulsions. Therefore, a deep understanding of structure–property dependences would allow for the design of the mixtures of rhamnolipids with properties adjusted for different applications. This could be reached by applying quantitative structure–property relationship (QSPR) models. The QSPR is nowadays one of the most commonly used *in silico* method for testing chemicals [34]. It is a tool for computational modelling, which shows molecular structures of tested chemicals in terms of descriptors later used to predict physical and biological properties [35].

To determine physical or biological activity quantitatively, minimal concentration of tested chemical giving some response is used. By analysis of those responses we can evaluate parameters having the highest influence on the activity of tested compounds. To obtain such information physicochemical parameters (descriptors) are used in QSPR models. Descriptors characterize tested chemical and should be selected by taking into account their relevance for the property under investigation. Different features of the molecule, such as thermodynamic, structural, and electron, can be described by descriptors. QSPR analysis allows then for the determination of most significant descriptors. In addition, it allows for the prediction of the properties of investigated compound within a certain descriptor range.

Even though methods of QSPR for surfactants [36] and ionic liquids are common [37], surprisingly no one has yet attempted to predict the MSR. This is especially unusual, considering the industrial importance of the parameter. Commonly, in technological applications relying on incorporation of insoluble drugs or active anti-aging substances into micelles or other carriers, the general rule is the larger the MSR, the better. However, this is not always the case; for example, a lower MSR opens the possibility to separate components in biological matrices or model non-aqueous phase liquids of environmental importance [38], as shown in Table 1.

**Table 1.** Exemplary results of selective solubilization of components from matrices of biological or environmental significance with an application of different surfactants. Shown are compounds selectively solubilized over other components of the matrix, thereby highlighting the need for a targeted molar solubilization ratio (MSR).

Process	Surfactant	Matrix	Compound Selectively Solubilized	Ref.
Separation of proteins with similar, low molecular size from their ternary mixture	<i>Anionic surfactant:</i> Aerosol OT/AOT (sodium di-2-ethylhexyl sulfosuccinate)	Ternary mixture of LMW proteins (ribonuclease-a, cytochrome-c and lysozyme)	Ribonuclease-a or cytochrome-c or lysozyme	[39]
Selective solubilization	<i>Nonionic surfactants:</i> Brij 30, C <sub>12</sub> E <sub>4</sub> Brij 35, C <sub>12</sub> E <sub>23</sub> Tergitol NP10, C <sub>9</sub> PE <sub>10.5</sub> Triton X-100, C <sub>8</sub> PE <sub>9.5</sub> Tween 80, C <sub>18</sub> S <sub>6</sub> E <sub>20</sub>	Naphthalene, hexadecane and/or phenantrene	Naphtalene or phenantrene	[40]
Selective solubilization	<i>Cationic surfactants:</i> Cetylpyridinium chloride, C <sub>16</sub> PyrCl Dodecyl ammonium chloride, C <sub>12</sub> NH <sub>3</sub> Cl Dodecyl trimethyl ammonium chloride, N[C <sub>12</sub> ,C,C,C]Cl <i>Anionic surfactants:</i> Sodium dodecyl sulfate, C <sub>12</sub> SO <sub>4</sub> Na Sodium diamyl sulfosuccinate (Aerosol AY)	Benzene-hexane mixtures	Benzene	[41]
Selective solubilization of PAHs from NAPL	<i>Nonionic surfactants:</i> Brij 30, C <sub>12</sub> E <sub>4</sub> Brij 35, C <sub>12</sub> E <sub>23</sub> Tergitol NP10, C <sub>9</sub> PE <sub>10.5</sub> Triton X-100, C <sub>8</sub> PE <sub>9.5</sub> Tween 80, C <sub>18</sub> S <sub>6</sub> E <sub>20</sub>	Ternary mixtures of naphthalene, phenantrene and hexadecane	Phenantrene <sup>1</sup>	[42]

Table 1. Cont.

Process	Surfactant	Matrix	Compound Selectively Solubilized	Ref.
Competitive solubilization of naphthalene and pyrene	<i>Nonionic surfactants:</i> Brij30, C <sub>12</sub> E <sub>4</sub> Brij56, C <sub>16</sub> E <sub>10</sub> <i>Cationic surfactants:</i> Dodecylethyltrimethyl ammonium bromide, N[C <sub>12</sub> ,C <sub>2</sub> ,C,C]Br Cethyltrimethyl ammonium chloride, N[C <sub>16</sub> ,C,C,C]Br <i>Anionic surfactant:</i> Sodium dodecyl sulfate, C <sub>12</sub> SO <sub>4</sub> Na	Binary mixtures of pyrene and naphthalene	Pyrene	[38]
Selective solubilization of glycerolipid and glycoprotein species from isolated human erythrocyte membranes	<i>Nonionic surfactant:</i> Triton X-100	Isolated human erythrocyte membranes	Glycolipid, glycerolipid	[43]

<sup>1</sup> or reverse, depending on the type of surfactant used in experiment; E—oxyethylene group; PE—oxypropylene group; types of surfactants were marked in italic.

Biosurfactants have a wide range of characteristics, with a wide range of MSR as shown in Table 2. The question is, therefore, which biosurfactant is most suitable for which application? To answer that question, we need to be able to predict the MSR. Even though methods for predicting micelles formation are common [36,44], prediction of the MSR has never been attempted. The only literature example of a related concept concerns microemulsions. The hydrophilic–lipophilic deviation (HLD) concept [45,46] was used to determine the composition of surfactant formulations for cosmetics, drug delivery and detergency [47], oil spills removal [48], enhanced oil recovery [49], also considering biosurfactant application [47]. However, the concept uses several descriptors that have to be determined experimentally prior to calculation, i.e., optimal salinity, critical curvature, equivalent carbon atoms number and empirical constant K. In addition, the HLD concept concerns Winsor type III microemulsion. This middle phase, bicontinuous system is known for its high solubilization capacity, but can be formed only in a narrow salinity [50,51], and temperature range [47]. That excludes many practical applications of the microemulsion, as even simple dilution of the formulation for cleaning purposes would disturb the equilibrium of the system and lead to transformation to Winsor type I or type II.

**Table 2.** Range and diversity of different parameters used for creation of model to predict MSR of BS in aqueous system. These represent the bounding box applicability domain according to Sahigara et al. [52].

Parameter	Impurity	logP <sub>BS</sub>	CMC [mg/L]	logP <sub>sol</sub>	V <sub>m sol</sub> . [Å <sup>3</sup> ]	pH	T [°C]	MSR
Min. value	0	4.51	0.0010	2.69	94.19	4	20	1.35 × 10 <sup>−5</sup>
Max. value	2	5.21	700.00	9.00	314.59	9	30	82.10
Mean	0.73	4.88	101.98	5.80	207.04	6.76	24.0	4.97
SD	0.80	0.22	118.20	1.99	59.42	0.78	2.7	16.06

Therefore, the purpose of this research was to use a series of chemical and phenomenological descriptors to predict the solubilization efficiency of different compounds in biosurfactant solutions, and then rank them upon their significance. The prediction will use the descriptors of both, the biosurfactant, and the solubilizate, but also the conditions of the measurement, which to our knowledge, is a first such attempt. The data used for this come from literature [1,4,27,33,53–68] as well as experiments.

## 2. Results and Discussion

### 2.1. Solubilizing Properties of Rhamnolipid Biocomplex

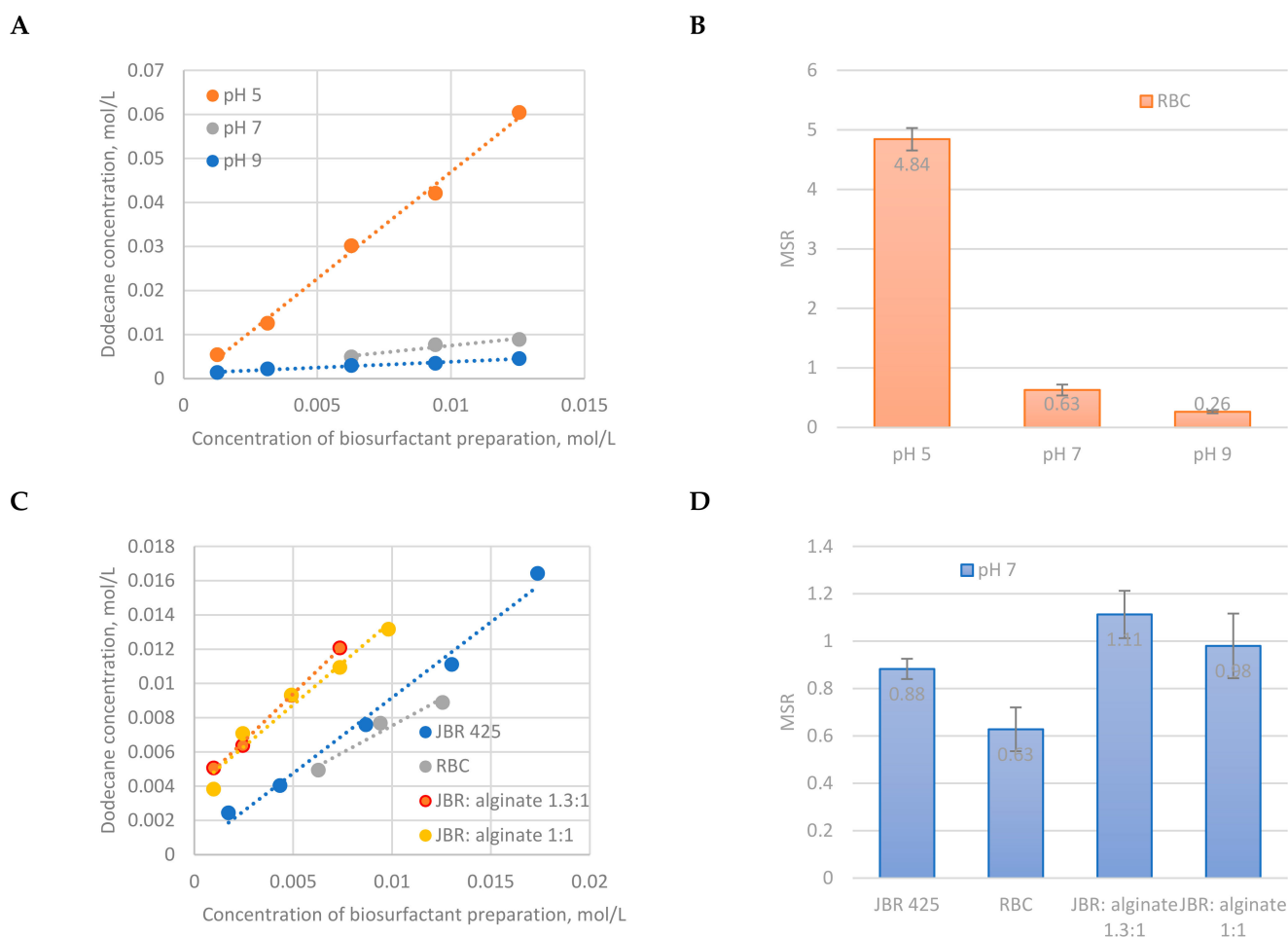
For QSPR models it is crucial to have a wide applicability domain, i.e., the space determined by descriptors on which the model was built and where it will be further applicable [69]. The literature dataset covered only narrow purity range of biosurfactants. In order to build a model covering a wider range of biosurfactants we therefore examined a new type of biosurfactant, i.e., rhamnolipid biocomplex with alginate (RBC). RBC belongs to 3rd class of purity according to classification published before [36], that was hardly represented in the literature dataset collected (ESI MSR dataset with MICE). What is more, rhamnolipid biocomplex is a cheaper and more environmentally friendly alternative than purified rhamnolipids [36].

We have determined the influence of pH on the solubilization efficiency of dodecane in RBC solutions (Figure 1A). We observed that the concentration of dodecane increased with increasing RBC concentration. This is due to parallel increase of number of micelles to receive solubilize. The increase was very steep at pH 5 and significantly milder at pH 7 and 9. The change of slope was obviously reflected by the decrease of MSR values from  $4.84 \pm 0.19$  in acidic to  $0.26 \pm 0.02$  in alkaline conditions (Figure 1B). This can be explained by change of solubility and dissociation of RLs with change of pH. RLs are weak acids with  $pK_a$  of 5.6 [70]. At pH 5 RL-carboxylic groups are protonated and of limited solubility. We hypothesize that solubilize may then act as solubility nuclei that supports formation of micelles. Pre-micelles formation was previously described in literature [71], except in our case they would be filled with the solubilize. It was already reported that RLs are able to solubilize hydrocarbons at concentrations below the CMC [27]. With change of pH, surfactant aggregation behavior changes as well. Shin et al. [72] found that rhamnolipids form vesicles in acidic, and micelles in alkaline conditions. The latter have significantly lower solubilization capacity.

We have also performed solubilization of dodecane in the solutions of different biosurfactant preparations, namely with use of model RBC (i.e., pure RLs mixtures with alginate at different mass ratios) and pure RLs (JBR 425) (Figure 1C). The experiments were performed at pH 7, so that the influence of alginate on solubilization is not overshadowed by the impact of dissociation, solubility or pre-micellization of RLs, as discussed above. Surprisingly, no clear trend was observed. The highest MSR was found for model RBCs, i.e., JBR: Alginate mixtures ( $1.11 \pm 0.10$  and  $0.98 \pm 0.14$  for 1.3:1 and 1:1 weight ratio, respectively) as compared to  $0.88 \pm 0.04$  for pure RLs (Figure 1D). Therefore, one could conclude that the presence of alginate increases solubilizing properties of RLs although alginate is not surface active. However, the biocomplex had the lowest MSR ( $0.63 \pm 0.09$ ), although containing about 80% of RLs. It is possibly due to the presence of junk molecules. JBR is composed of pure RLs, and model biocomplexes are composed of JBR and alginate, whereas RBC is a product of biosynthesis that was not so deeply purified. It represents 2nd Impurity class as compared to JBR 425, which represents 0 Impurity class. It was already stated by us previously that junk molecules increase the CMC of RLs. At pH 7 the CMC of JBR was 41.5 mg/L, whereas it was 62.1 mg/L for RBC [36]. Junk molecules may weaken the solubilizing properties of RLs analogously to micellization properties. Increase of MSR observed for model RBCs (Figure 1D) would indicate change of interactions between RLs and alginate with increasing concentration of the latter one. Perhaps some adsorption at alginate interface or even saturation may occur, likewise in surfactant-protein systems [73]. A more detailed physicochemical research would be required to confirm that hypothesis.

The obtained MSR values for RBC, pure RLs, and their mixtures with alginate (model RBCs) at different pH conditions were further used to expand the applicability domain of the model to predict MSR in different biosurfactant systems (ESI MSR dataset).





**Figure 1.** Efficiency of dodecane solubilization in biosurfactant solutions: (A) The influence of pH on the dodecane concentration in rhamnolipid biocomplex solutions (RBC) and (B) corresponding molar solubilization ratios (MSRs) values; (C) dodecane concentration dependence on the concentration of different biosurfactant preparations at pH 7, and (D) corresponding MSRs. JBR represents pure RLs (JBR425 of Jeneil), mixtures of JBR and alginate represent model RBC. With increasing pH rhamnolipid aggregates change from vesicles to micelles and the sharp decrease of solubilization efficiency is observed for RBC solutions.

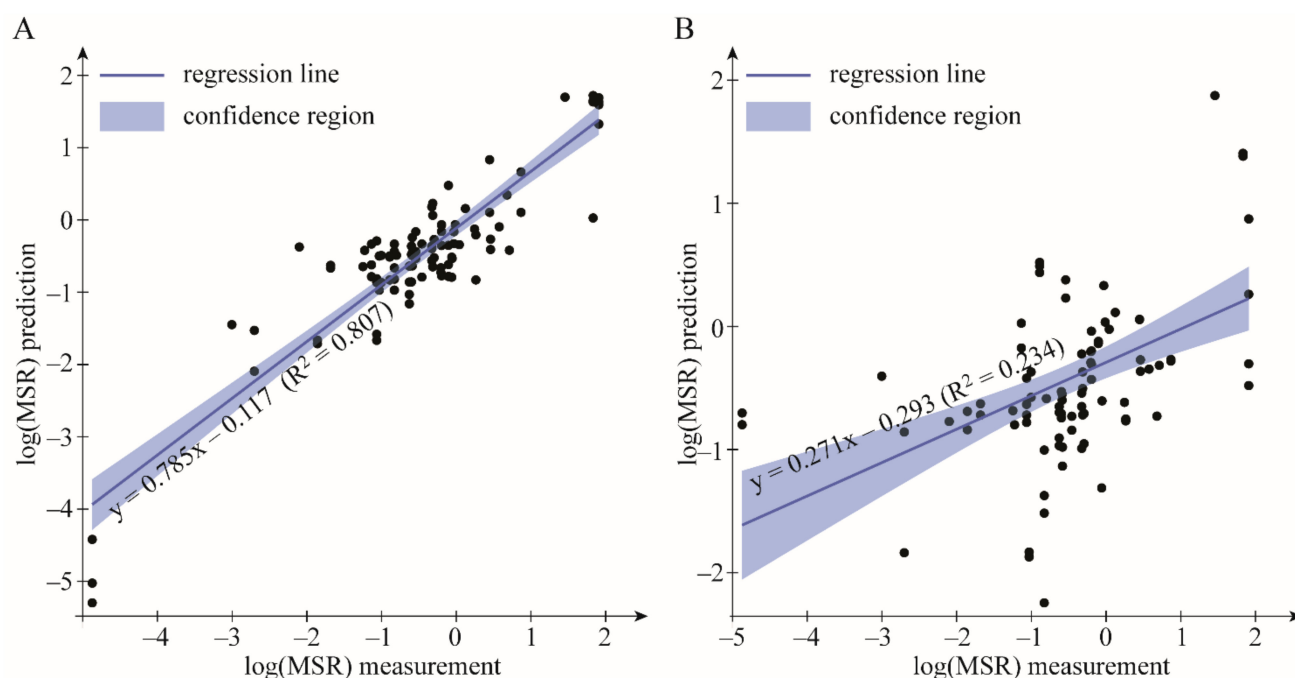
## 2.2. Predicting MSR with EA and KSVM

The model to predict MSR was built on the combined literature and experimental data and represented by the function  $\log\text{MSR} = f(\text{CMC}, \text{pH}, \log P_{\text{sol}}, V_{\text{m sol}}, T, (\text{im})\text{purity}, \log P_{\text{BS}})$  with evolutionary algorithm (EA) EureqaPro was generated 10 times. The results of the modelling are presented in Table 3 and Figure 2.

The external validation results showed an advantage of the EA over KSVM model. The average  $R^2_{\text{val}}$  for ten validation sets were equal to  $0.773 \pm 0.052$  and  $0.166 \pm 0.152$  for EA and KSVM, respectively (Table 3). The overall predictive power of EA model was also much higher ( $R^2 = 0.807$ , Figure 2A) than observed for KSVM model ( $R^2 = 0.234$ , Figure 2B). With  $R^2$  of 0.807 the correlation is considered to be strong [74]. Therefore, the MSR can be successfully predicted using EAs. To check if the numerical relationship is not by chance y-randomization technique was also applied. The method successfully validated the model developed with EA. As can be seen from the results of the y-randomization (Table 3), the real  $R^2$  was a statistical outlier, as opposed to the y-randomized  $R^2$  values, with a p value of 0.006, indicating that the relationship between the MSR and descriptors is non-random.

**Table 3.** Parameters obtained for ten models, used to obtain formula able to predict logMSR, after 115,000 generations. Average  $R^2_{\text{val}}$  and average  $\text{MSE}_{\text{val}}$  were calculated for ten validation sets. Also shown are results of the  $y$ -randomization, indicating that the relationships are not due to random chance.

Parameter	Chemometric Tool	
	EA	KSVM
External validation:		
$R^2_{\text{val}} \pm \text{CI}$	$0.773 \pm 0.052$	$0.166 \pm 0.152$
$\text{MSE}_{\text{val}} \pm \text{CI}$	$0.296 \pm 0.084$	$1.176 \pm 0.483$
Y-randomization:		
G	2.613	-
G (Critical value)	2.355	-
$p$ -value	0.006	-



**Figure 2.** The estimated predictive power of the model used for logMSR calculation. The opaque region is 95% confidence interval (CI) of the regression. As can be seen the evolutionary algorithm (EA) approach performs better (A) than the kernel support vector machines (KSVM) (B). Variable  $x$  in regression line equations denotes the log(MSR) measurement. The  $R^2$  values displayed in the figures are given with respect to outcomes of all ten testing sets.

In predicting logMSR kernel support vector machines (KSVM) did not perform satisfactorily ( $R^2_{\text{val}} = 0.166 \pm 0.152$ ,  $R^2 = 0.234$ ). Although, KSVM is considered an effective regression tool for small datasets, it still suffers from the curse of dimensionality [75]. Although it handles small datasets better than other regression techniques it still may become useless if the ratio of observations to features is not big enough. The genetic algorithm approach, like the one presented by Eureka, is much more likely to succeed in case of data sparsely distributed in the feature space. If the amount of the data does not allow to estimate the functional relationship with statistical methods, fitting a surface may still provide a tool for prediction.

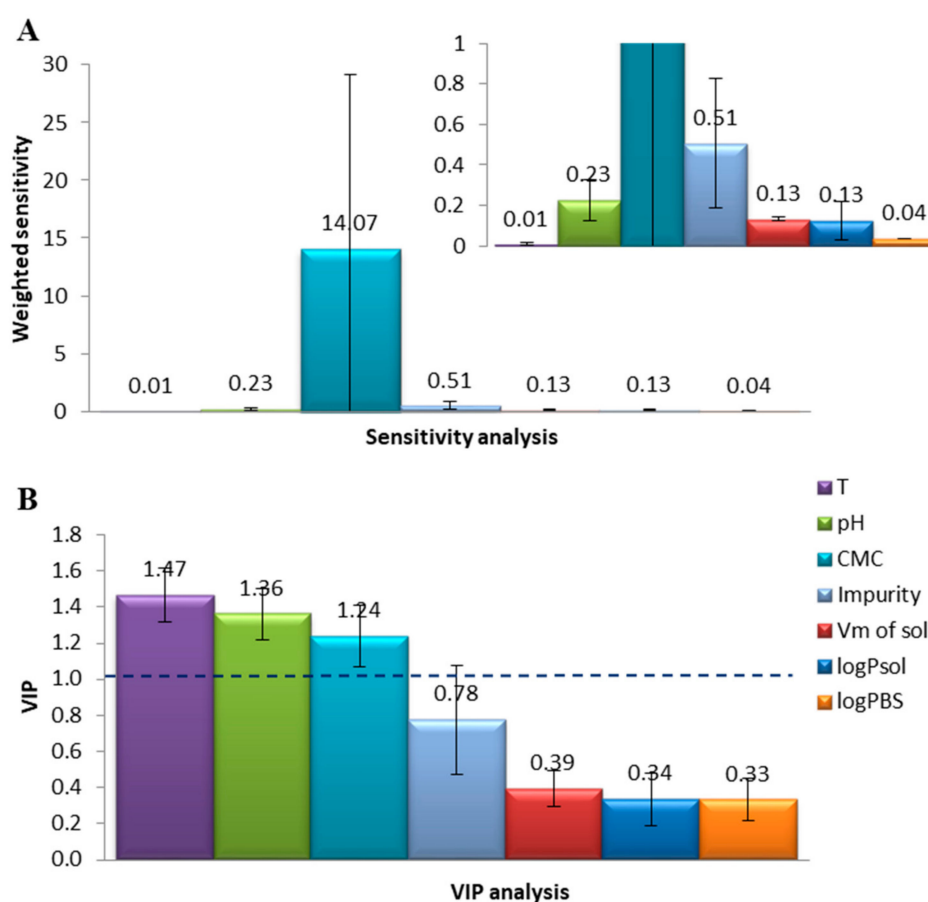
Support vector machines are known to be useful in case of small datasets as it takes into consideration the points that support the margin [76]. However, the dataset has to be large enough to provide the sufficient number of points supporting the margin. Otherwise, it is impossible to build the function in kernel space that would reflect the real relationships

hidden in the data. If the dataset is unable to provide sufficient number of support vectors then the prediction becomes poor.

In the considered case, we are unable to build a reliable KSVM predictor for the given data, due to its insufficient size.

### 2.3. Sensitivity Analysis

All ten models (given in Table S5) created with EA for the dataset were subjected to the sensitivity analysis [36]. The results of the analysis are shown in Figure 3A,B. As can be seen from Figure 3A, according to EA sensitivity analysis MSR strongly depends on CMC, and the influence of other parameters remains negligible. The importance of CMC was confirmed by PLS analysis (VIP  $\geq 1$  [77]). Here in turn, also pH and T were found to significantly alter solubilization efficiency (Figure 3B).



**Figure 3.** Sensitivity analysis for modelling of logMSR in biosurfactant solutions: (A) Average sensitivity towards each of descriptors weighted according to number of appearances, obtained with evolutionary algorithm (EA); (B) average variable importance in projection (VIP) analysis performed with PLS regression method. Error bars represent 95% CI. The Eureka evolutionary algorithm (EA) provides random model equations. The error in the sensitivity of the CMC is due to the large range from 0.78 to 158 as shown in Table S5.

The results of sensitivity analysis are in agreement with literature. The dependence of the MSR on CMC is evident from Equation (1), as micellar solubilization requires the presence of micelles. As CMC is sensitive to pH changes, therefore this environmental factor appears to alter the overall solubilization efficiency. Further, temperature may also affect MSR. Just as in case of pH, this happens indirectly—by influencing CMC that already appeared crucial for MSR prediction within both sensitivity analyzes (Figure 3A,B).

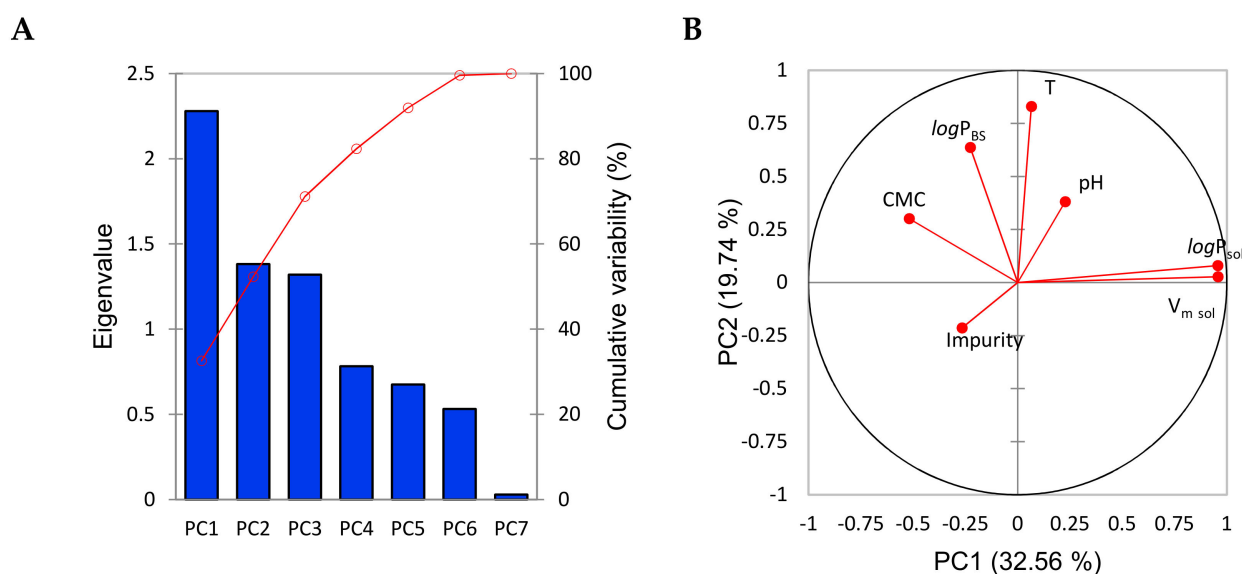
Surprisingly, the nature of solubilize, namely its hydrophobicity/hydrophilicity (expressed here as logP<sub>sol</sub>) seems to have no apparent effect on solubilization. Although lit-



erature concerning micellar solubilization indicates that strongly hydrophobic solubilizes (e.g., alkanes) incorporate in the core of micelle, and hydrophilic solubilizes incorporate in palisade layer [78,79], this effect was found negligible for predicting MSR by both VIP and EA sensitivity analyzes (Figure 3A,B).

#### 2.4. PCA and Descriptor Overlap

PCA is a technique that projects the data into another feature space in which the arising features are uncorrelated. The data features in the new space are called principal components (PCs). The PCs are ordered due to the decreasing amount of information (Eigenvalue) introduced into the dataset. They are usually presented in the form of a bar chart (Figure 4A). Such knowledge allows to reduce the data dimensionality as the vectors positioned at the last places usually introduce small amounts of information and can be neglected without the significant losses. Apart from dimensionality reduction, PCA can be used to estimate the correlation relationships of the original features. For this purpose projection of the original features onto the plane created by the two most significant PCs is done. The graphical presentation of this is the correlation plot (Figure 4B). In such a plot vectors pointing along the similar lines are correlated. If they point at the same direction the correlation is positive, if opposite the correlation is negative. As can be seen on Figure 4B, both  $V_m$  and  $\log P$  of the solubilize are strongly overlapping, indicating that they provide similar information to the model. The similarity of  $\log P_{sol}$  and  $V_{m\ sol}$  is clear, as a larger  $V_m\ sol$  results in a larger  $\log P_{sol}$ , considering the collected range of the solubilizes. The PCA results are in accordance with the results provided by EA sensitivity analysis. None of 10 EA models delivered used these two descriptors simultaneously (Table S5). Other variables are uncorrelated with any of the other features.



**Figure 4.** The results of the PCA analysis are shown. (A) shows the scree plot of the eigen-values, indicating that the data can be reduced to about six variables. In (B), correlation plot of descriptors is presented.

### 3. Materials and Methods

#### 3.1. Materials

Rhamnolipid biocomplex (RBC) was biosynthesized by *Pseudomonas* sp. PS-17 as described previously [36] and is a complex of RLs with an exopolysaccharide, alginate. JBR 425 (lot. No. 040714) in a form of the 25% neutral aqueous solution of dirhamnolipid and monorhamnolipid at 0.97:1 (*w/w*) ratio was kindly provided by the Jeneil Biosurfactant Company (Saukville, WI, USA). This sample is referred to pure RLs. Sodium alginate (W201502) was purchased in Sigma-Aldrich (Poznań, Poland). Dodecane (99+, Lancaster)



was purified from polar additives by passing through a column filled with calcinated  $\text{Al}_2\text{O}_3$  (120 °C, 2 h). Isopropanol (p.a., POCh Gliwice, Poland) was used as a solvent for GC analysis.

### 3.2. Dodecane Solubilization

Rhamnolipid biocomplex (RBC), pure RLs (JBR425) and their mixtures with alginate at 1.3:1 and 1:1 weight ratio (referred to model RBCs) were used as biosurfactant preparations of different purity for the current research. Aqueous solutions of the biosurfactant preparations at the following concentrations 1.0, 2.5, 5.0, 7.5, and 10.0 g/L were prepared by series dilution of 10 g/L stock solution. Ultrapure water (D4700, Barnstead) was used for preparing the solutions. NaOH and HCl were used to set the pH. Aqueous (5 mL) and oil phases (0.5 mL of dodecane) were mixed in 10 mL vials at 1500 rpm for 24 h with a horizontal shaker (IkaVibrax VXR, Ika Technologies, Boutersem, Belgium) at room temperature. Unsolubilized dodecane (excess) was separated from micellar phase by centrifugation at 5000 rpm for 15 min (MPW-350R, MPW Med. Instruments, Warsaw, Poland). Subsequently, clear micellar phase was separated with a syringe through a silicone sealed hole in a bottom of the vial and analyzed for the content of dodecane. All experiments were run in triplicate.

### 3.3. Solubilization Efficiency

Gas chromatography was used for determination of dodecane solubilization efficiency. One milliliter of micellar phase was diluted with 4 mL of 2-propanol. Chrompack CP 9001 with XTI-5 capillary column (30 m × 0.32 mm, 1 μm) was used for GC analysis. The analysis was performed at 180 °C, using a split feeder and FID at 250 °C. The injection volume was 1 μL. An external calibration was used for quantitative analysis.

A simplified equation 1 was used to calculate MSR values. As the solubility of dodecane is negligible,  $C_{CMC} \sim 0$  was assumed. Considering high surfactants concentrations used as compared to CMC, and linearity of  $C = f(C_{surf})$  function in the examined range of surfactant concentrations, the MSR equals to the slope (i.e., directional factor  $a$ ) as follows:

$$MSR = \frac{C - C_{CMC}}{C_{surf} - C_{CMC}} = \frac{C}{C_{surf}} = \frac{\Delta C}{\Delta C_{surf}} = tg\alpha = a \quad (2)$$

where  $\alpha$  is a slope of  $C = f(C_{surf})$  function, and  $a$  is a directional factor of thereof.

### 3.4. Data Collection

The modelling was performed on a dataset of MSR values collected from literature and experimental data obtained for the purpose of this research. Google Scholar, which provides a high coverage [80], was used for literature data collection. The years 1980–2017 were covered in the search, however the first article was found in 1994. The following keywords were used to search for the MSR data: Biosurfactant, rhamnolipid, monorhamnolipid, dirhamnolipid, sophorolipid, emulsan, lichenisyn, surfactin, solubilization, MSR, and WSR (molar and weight solubilization ratio, respectively). As a result, 21 publications were found, resulting in 44 unique data points. However, only 11 of them provided continuous data needed for modelling, resulting in a dataset with 23 literature data points. Missing data in remaining 16 discontinuous datapoints were filled with Multiple Imputation by Chained Equations (MICE) technique. Finally, the dataset used for modelling consisted of 45 datapoints, out of which six were determined experimentally for the purpose of this research (Table 2 and ESI: MSR Dataset). WSR data were recalculated into MSR value using the Equation (S1). CMC was taken from each paper. Structural descriptors of solubilizate and rhamnolipid (molecular volume  $V_m$  and logarithm of octanol-water partition coefficient  $\log P$ ) were calculated using Molinspiration [81]. If the rhamnolipid was present as a mixture, the weight percent mean was used, as shown in Tables S2 and S3. The state of biosurfactant purification was expressed numerically based on reversed purity scale published before [36] and is presented in Table S4.

### 3.5. MICE

Multiple Imputation by Chained Equations (MICE) was used to extend the dataset (composed initially of 29 data points), and fill the missing data gaps. Here we used an implementation of MICE within XLSTAT (Adinsoft) [82]. Predictive mean matching was used with 14 iterations. MICE methodology was validated as described by Łozińska et al. [77]. The normalized mean root squared error (NMRSE) graph is shown in Supplementary Figure S1. MICE was utilized here to show that the model can potentially also be applied to a larger variety of biosurfactants (i.e., alasan, and flavolipids).

### 3.6. Computational Modelling

The dataset was randomly divided into training and validation sets at 8:2 ratio [37,83]. This allows for external validation of the model [84]. In total 10 random training and validation sets were created in order to obtain a truly representative relationship. This methodology has been described in detail by [37,85].

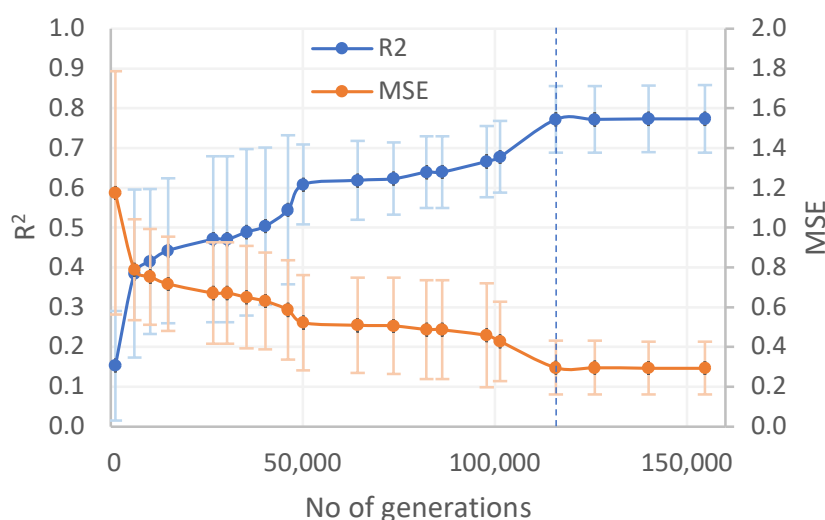
Two methods for constructing models were applied. One was a parametric regression method in which an expression is fitted to the provided data. The expression was constructed by evolutionary algorithm (EA). For this purpose EureqaPro software was used (v1.24.0, build 9367). As the second kernel support vector machine (KSVM) regression was performed. For this purpose, kernlab package in R [86] was used.

In total seven molecular descriptors related to MSR were used to develop a model. Those included physical (CMC) and chemical descriptors (logP of BS and solubilize, impurity of BS), molecular volumes ( $V_{m\ sol}$ ) and conditions present in aqueous system (pH, T). The models were searched in a form of the following function  $\log MSR = f(\text{CMC}, \log P_{BS}, (\text{im})\text{purity}, \log P_{sol}, V_{m\ sol}, \text{pH}, T)$ .

Detailed descriptions of EA, PLS, and KSVM techniques have previously been provided by us elsewhere [36,77].

### 3.7. EA Generations

Due to the fact, that EureqaPro generates different equations for the same data sets each instance it is run, modelling was repeated ten times to assure high reliability of final model (the methodology for this was discussed further in our previous publication [36]). Based on those ten measurements average  $R^2$  and average mean squared error (MSE) were calculated to determine the end point of modelling (Figure 5).



**Figure 5.** Determination of the modeling endpoint by the analysis of mean  $R^2$  and mean MSE with SD. The dashed line indicates the endpoint at 115,000 generations. The endpoint was chosen due to the plateau region and lower SD in MSE and  $R^2$ .

### 3.8. Model Validation—Y-Randomization

Y-randomization was performed to exclude an accidental relationship between the descriptors, and the dependent variable (MSR). In this case, 10 randomizations of the MSR were conducted. The descriptors were left unchanged. Each set was modelled separately using the EA as described before. The resulting  $R^2$  was compared to the non-randomized  $R^2$  of the original set, and the Grubbs outlier test was used to determine the  $p$ -value of the outlier.

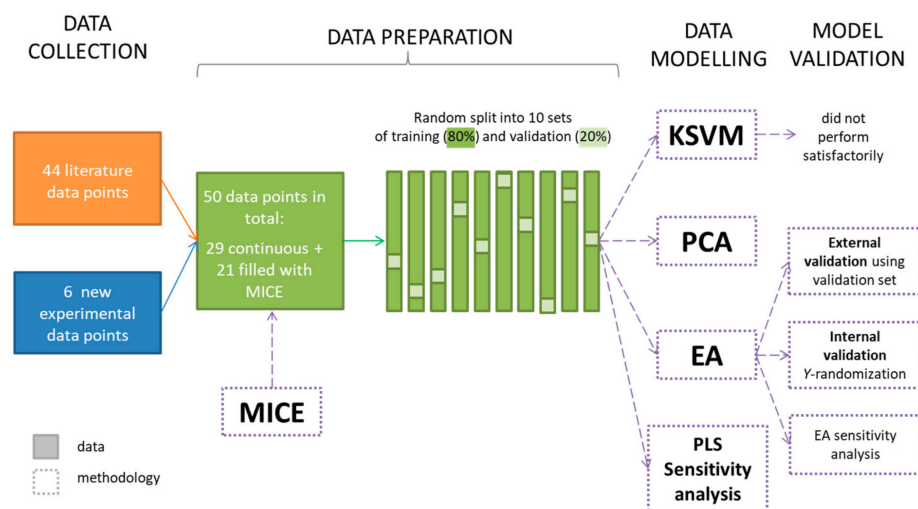
### 3.9. SVM Regression Methodology

In order to perform KSVM regression each of the data features was rescaled to the range of values varying from  $-1$  to  $1$ . A Gaussian kernel was chosen as this is usually the best one for non-linear data [87]. The set of optimal hyper-parameters was determined using the grid search with leave one out cross validation. The error margin tolerance was set to  $\varepsilon = 0.01$ .

### 3.10. Sensitivity for EA and KSVM

The sensitivity of the model towards descriptors was analyzed only in case of the EA outcome, as KSVM did not provide satisfactory results. Such sensitivity analysis was performed with two methods, namely using EA sensitivity analysis [36] and partial least squares variable importance of projection (PLS VIP) analysis [77,88]. The XLSTAT implementation of PLS regression was used [82,89]. For cross validation, the jackknife leave one out method was used [90]. The variables importance for the projection VIPs were determined according to [91]. PCA was also conducted within XLSTAT.

An overall schematic of the research procedure is shown in Figure 6.



**Figure 6.** Schematic representation of the procedure used to create the models. EA — evolutionary algorithm; KSVM—kernel support vector machine; MICE—multiple implementation by chain equations; PCA—principal component analysis; PLS—partial least squares.

## 4. Conclusions

To our knowledge the prediction of solubilization efficiency (MSR) in RLs solutions was described for the first time. We have shown that MSR can be successfully predicted with EA ( $R^2_{\text{val}} = 0.773 \pm 0.052$ ). As opposed to HLD concept [48,49], vast majority of descriptors for our MSR approach can be easily calculated with a software widely available online. Therefore, the experimental effort is minimized.

Additionally, we have ranked the parameters influencing the MSR upon their significance. This in turn may enhance the ease of designing experimental work with rhamnolipids, as the most crucial parameters for rhamnolipid solubilizing properties were determined and systematized.

The model will allow biosurfactants to be designed for specific applications, which will enhance the use of these green surface active molecules. What is more, we aimed at changing the approach of surfactant application. Currently, research is often aimed to determine the best application for a surfactant that was either selected or bio/synthesized for the purpose of research [92–95]. However, here we propose to determine the optimal surfactant a priori for the application of interest, e.g., to pre-select best rhamnolipid biosurfactant for removing oil spill of known composition. We propose here the following hypothetical scenario, where we assume that on one spring day ( $T = 20\text{ }^{\circ}\text{C}$ ) a *p*-xylene ( $\log P = 2.83$ ,  $V_m = 117.17\text{ \AA}^3$  [81]) tank truck is damaged due to road incident and the chemical spills over the road and neighboring acidic ground (pH 5), which raises an urgent need for ground remediation. The MSR calculator that we provide (ESI MSR calculator) allows for quick selection of a biosurfactant that can be used, indicating the optimal biosurfactant will be found among alasan (calculated MSR = 19.0), monoRL mixture (calculated MSR = 7.3), or mixture of JBR425 and alginate 1:1 (calculated MSR = 2.34). Combining our new approach with the required parameters of biosynthesis would result in “a recipe for rhamnolipid” for a targeted application, which we are currently investigating further. The MSR model is limited by range of parameters as defined by the applicability domain that comprise the model. It is feasible, for example, that biosurfactants can be developed that have larger or smaller MSR values in the future. The model should then be updated.

**Supplementary Materials:** The following are available online, complete dataset of MSRs used for prediction, EA equations, and MSR calculator.

**Author Contributions:** Conceptualization, I.E.K.-C. and C.J.; experimental methodology, I.E.K.-C., E.H., and K.M.; computational methodology, I.E.K.-C., W.A., and C.J.; investigation, I.E.K.-C., W.A., and C.J.; resources, E.K.; data collection and curation, I.E.K.-C. and A.K.-S.; writing—original draft preparation, I.E.K.-C., W.A., and C.J.; funding acquisition, I.E.K.-C.; writing—review and editing, A.M., A.G.-R., and P.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Narodowe Centrum Nauki, Poland (2017/01/X/ST5/02077).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in electronic supplementary information [ESI, ESI MSR dataset with MICE, ESI MSR calculator].

**Acknowledgments:** Authors kindly acknowledge Kamila Kucińska, Krystyna Rybicka and Małgorzata Borecka for their technical support, and Joanna Panek ([www.joannapanek.com](http://www.joannapanek.com)) for the creation of graphical abstract.

**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 data; in the writing of the manuscript, or in the decision to publish the results.

**Sample Availability:** Samples of the RBC are available from the authors.

## Abbreviations

CI, confidence interval; CMC, critical micellar concentration; EA, evolutionary algorithm; KSVM, kernel support vector machines;  $\log P$ , logarithm of octanol-water partition coefficient; MSE, mean square error; MSR, molar solubilization ratio; NRMSE, normalized root mean squared error; PC, principal component; PCA, principle component analysis; QSPR, quantitative structure–property relationship; RBC, rhamnolipid biocomplex with alginate; SD, standard deviation;  $V_m$ , molecular volume.

## References

1. Clifford, J.S.; Ioannidis, M.A.; Legge, R.L. Enhanced aqueous solubilization of tetrachloroethylene by a rhamnolipid biosurfactant. *J. Colloid Interface Sci.* **2007**, *305*, 361–365. [[CrossRef](#)] [[PubMed](#)]
2. Whang, L.-M.; Liu, P.-W.G.; Ma, C.-C.; Cheng, S.-S. Application of biosurfactants, rhamnolipid, and surfactin, for enhanced biodegradation of diesel-contaminated water and soil. *J. Hazard. Mater.* **2008**, *151*, 155–163. [[CrossRef](#)] [[PubMed](#)]

3. Mulligan, C.N.; Yong, R.; Gibbs, B. Surfactant-enhanced remediation of contaminated soil: A review. *Eng. Geol.* **2001**, *60*, 371–380. [[CrossRef](#)]
4. Liu, Y.; Zeng, G.; Zhong, H.; Wang, Z.; Liu, Z.; Cheng, M.; Liu, G.; Yang, X.; Liu, S. Effect of rhamnolipid solubilization on hexadecane bioavailability: Enhancement or reduction? *J. Hazard. Mater.* **2017**, *322*, 394–401. [[CrossRef](#)] [[PubMed](#)]
5. Nasongkla, N.; Bey, E.; Ren, J.; Ai, H.; Khemtong, C.; Guthi, J.S.; Chin, S.-F.; Sherry, A.D.; Boothman, D.A.; Gao, J. Multifunctional polymeric micelles as cancer-targeted, MRI-ultrasensitive drug delivery systems. *Nano Lett.* **2006**, *6*, 2427–2430. [[CrossRef](#)] [[PubMed](#)]
6. Kaur, P.; Garg, T.; Rath, G.; Murthy, R.; Goyal, A.K. Surfactant-based drug delivery systems for treating drug-resistant lung cancer. *Drug Deliv.* **2016**, *23*, 717–728. [[CrossRef](#)]
7. Maldonado-Valderrama, J.; Wilde, P.; Macierzanka, A.; Mackie, A. The role of bile salts in digestion. *Adv. Colloid Interface Sci.* **2011**, *165*, 36–46. [[CrossRef](#)]
8. Scott, M.J.; Jones, M.N. The biodegradation of surfactants in the environment. *Biochim. Biophys. Acta Biomembr.* **2000**, *1508*, 235–251. [[CrossRef](#)]
9. Tehrani-Bagha, A.R.; Oskarsson, H.; Van Ginkel, C.; Holmberg, K. Cationic ester-containing gemini surfactants: Chemical hydrolysis and biodegradation. *J. Colloid Interface Sci.* **2007**, *312*, 444–452. [[CrossRef](#)]
10. Toribio, J.; Escalante, A.E.; Soberón-Chávez, G. Rhamnolipids: Production in bacteria other than *Pseudomonas aeruginosa*. *Eur. J. Lipid Sci. Technol.* **2010**, *112*, 1082–1087. [[CrossRef](#)]
11. Nitschke, M.; Costa, S. Biosurfactants in food industry. *Trends Food Sci. Technol.* **2007**, *18*, 252–259. [[CrossRef](#)]
12. Chrzanowski, Ł.; Ławniczak, Ł.; Czaczyk, K. Why do microorganisms produce rhamnolipids? *World J. Microbiol. Biotechnol.* **2012**, *28*, 401–419. [[CrossRef](#)] [[PubMed](#)]
13. Abdel-Mawgoud, A.M.; Lépine, F.; Déziel, E. Rhamnolipids: Diversity of structures, microbial origins and roles. *Appl. Microbiol. Biotechnol.* **2010**, *86*, 1323–1336. [[CrossRef](#)] [[PubMed](#)]
14. Hošková, M.; Schreiberová, O.; Ježdík, R.; Chudoba, J.; Masák, J.; Sigler, K.; Řezanka, T. Characterization of rhamnolipids produced by non-pathogenic *Acinetobacter* and *Enterobacter* bacteria. *Bioresour. Technol.* **2013**, *130*, 510–516. [[CrossRef](#)] [[PubMed](#)]
15. Elshikh, M.; Funston, S.; Chebbi, A.; Ahmed, S.; Marchant, R.; Banat, I.M. Rhamnolipids from non-pathogenic *Burkholderia thailandensis* E264: Physicochemical characterization, antimicrobial and antibiofilm efficacy against oral hygiene related pathogens. *New Biotechnol.* **2017**, *36*, 26–36. [[CrossRef](#)] [[PubMed](#)]
16. Wigneswaran, V.; Nielsen, K.F.; Sternberg, C.; Jensen, P.R.; Folkesson, A.; Jelsbak, L. Biofilm as a production platform for heterologous production of rhamnolipids by the non-pathogenic strain *Pseudomonas putida* KT2440. *Microb. Cell Fact.* **2016**, *15*, 181. [[CrossRef](#)] [[PubMed](#)]
17. Twigg, M.S.; Tripathi, L.; Zompra, A.; Salek, K.; Irerere, V.; Gutierrez, T.; Spyroulias, G.; Marchant, R.; Banat, I.M. Identification and characterisation of short chain rhamnolipid production in a previously uninvestigated, non-pathogenic marine pseudomonad. *Appl. Microbiol. Biotechnol.* **2018**, *102*, 8537–8549. [[CrossRef](#)]
18. Özdemir, G.; Peker, S.; Helvacı, S. Effect of pH on the surface and interfacial behavior of rhamnolipids R1 and R2. *Colloids Surf. A* **2004**, *234*, 135–143. [[CrossRef](#)]
19. Jemil, N.; Ayed, H.B.; Hmidet, N.; Nasri, M. Characterization and properties of biosurfactants produced by a newly isolated strain *Bacillus methylotrophicus* DCS1 and their applications in enhancing solubility of hydrocarbon. *World J. Microbiol. Biotechnol.* **2016**, *32*, 175. [[CrossRef](#)]
20. Kłosowska-Chomiczewska, I.E.; Medrzycka, K.; Karpenko, E. Biosurfactants—Biodegradability, toxicity, efficiency in comparison with synthetic surfactants. In Proceedings of the Research and Application of New Technologies in Wastewater Treatment and Municipal Solid Waste Disposal in Ukraine, Sweden, and Poland, Cracow, Poland, 17–19 October 2011; pp. 1–9.
21. Li, G.; Lan, G.; Liu, Y.; Chen, C.; Lei, L.; Du, J.; Lu, Y.; Li, Q.; Du, G.; Zhang, J. Evaluation of biodegradability and biotoxicity of surfactants in soil. *RSC Adv.* **2017**, *7*, 31018–31026. [[CrossRef](#)]
22. Özdemir, G.; Sezgin, Ö. Keratin–rhamnolipids and keratin–sodium dodecyl sulfate interactions at the air/water interface. *Colloids Surf. B* **2006**, *52*, 1–7. [[CrossRef](#)] [[PubMed](#)]
23. Ju, L.-K.; Dashtbozorg, S.; Vongpanish, N. Wound Dressings with Enhanced Gas Permeation and Other Beneficial Properties. U.S. Patent 9,468,700, 18 October 2016.
24. Ławniczak, Ł.; Marecik, R.; Chrzanowski, Ł. Contributions of biosurfactants to natural or induced bioremediation. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 2327–2339. [[CrossRef](#)] [[PubMed](#)]
25. Chrzanowski, Ł.; Dziadas, M.; Ławniczak, Ł.; Cyplik, P.; Białas, W.; Szulc, A.; Lisiecki, P.; Jeleń, H. Biodegradation of rhamnolipids in liquid cultures: Effect of biosurfactant dissipation on diesel fuel/B20 blend biodegradation efficiency and bacterial community composition. *Bioresour. Technol.* **2012**, *111*, 328–335. [[CrossRef](#)] [[PubMed](#)]
26. Marecik, R.; Wojtera-Kwiczor, J.; Ławniczak, Ł.; Cyplik, P.; Szulc, A.; Piotrowska-Cyplik, A.; Chrzanowski, Ł. Rhamnolipids increase the phytotoxicity of diesel oil towards four common plant species in a terrestrial environment. *Water Air Soil Pollut.* **2012**, *223*, 4275–4282. [[CrossRef](#)]
27. Zhong, H.; Yang, X.; Tan, F.; Brusseau, M.L.; Yang, L.; Liu, Z.; Zeng, G.; Yuan, X. Aggregate-based sub-CMC solubilization of n-alkanes by monorhamnolipid biosurfactant. *New J. Chem.* **2016**, *40*, 2028–2035. [[CrossRef](#)]

28. Tiso, T.; Ihling, N.; Kubicki, S.; Biselli, A.; Schonhoff, A.; Bator, I.; Thies, S.; Karmainski, T.; Kruth, S.; Willenbrink, A.-L. Integration of genetic and process engineering for optimized rhamnolipid production using *Pseudomonas putida*. *Front. Bioeng. Biotechnol.* **2020**, *8*, 976. [\[CrossRef\]](#)
29. He, N.; Wu, T.; Jiang, J.; Long, X.; Shao, B.; Meng, Q. Toward high-efficiency production of biosurfactant rhamnolipids using sequential fed-batch fermentation based on a fill-and-draw strategy. *Colloids Surf. B* **2017**, *157*, 317–324. [\[CrossRef\]](#)
30. Jiang, J.; Zu, Y.; Li, X.; Meng, Q.; Long, X. Recent progress towards industrial rhamnolipids fermentation: Process optimization and foam control. *Bioresour. Technol.* **2020**, *298*, 122394. [\[CrossRef\]](#)
31. Moutinho, L.F.; Moura, F.R.; Silvestre, R.C.; Romão-Dumaresq, A.S. Microbial biosurfactants: A broad analysis of properties, applications, biosynthesis and techno-economical assessment of rhamnolipid production. *Biotechnol. Prog.* **2020**, e3093. [\[CrossRef\]](#)
32. Klosowska-Chomiczewska, I. Physicochemical Properties and Possibilities of Application of Biosurfactant Preparation Produced by *Pseudomonas PS-17* (pol.). Ph.D. Thesis, Gdańsk University of Technology, Gdańsk, Poland, 2014.
33. Zhang, Y.; Miller, R.M. Effect of rhamnolipid (biosurfactant) structure on solubilization and biodegradation of n-alkanes. *Appl. Environ. Microbiol.* **1995**, *61*, 2247–2251. [\[CrossRef\]](#)
34. Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C.D. Chem-bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.* **2002**, *102*, 783–812. [\[CrossRef\]](#)
35. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Klosowska-Chomiczewska, I.; Mędrzycka, K.; Hallmann, E.; Karpenko, E.; Pokynbroda, T.; Macierzanka, A.; Jungnickel, C. Rhamnolipid CMC prediction. *J. Colloid Interface Sci.* **2017**, *488*, 10–19. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Klosowska-Chomiczewska, I.E.; Artichowicz, W.; Preiss, U.; Jungnickel, C. Multicomponent ionic liquid CMC prediction. *Phys. Chem. Chem. Phys.* **2017**, *19*, 25309–25318. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Masrat, R.; Maswal, M.; Dar, A.A. Competitive solubilization of naphthalene and pyrene in various micellar systems. *J. Hazard. Mater.* **2013**, *244*, 662–670. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Goklen, K.E.; Hatton, T.A. Liquid-liquid extraction of low molecular-weight proteins by selective solubilization in reversed micelles. *Sep. Sci. Technol.* **1987**, *22*, 831–841. [\[CrossRef\]](#)
40. Hill, A.J.; Ghoshal, S. Micellar solubilization of naphthalene and phenanthrene from nonaqueous-phase liquids. *Environ. Sci. Technol.* **2002**, *36*, 3901–3907. [\[CrossRef\]](#)
41. Chaiko, M.A.; Nagarajan, R.; Ruckenstein, E. Solubilization of single-component and binary mixtures of hydrocarbons in aqueous micellar solutions. *J. Colloid Interface Sci.* **1984**, *99*, 168–182. [\[CrossRef\]](#)
42. Bernardes, L.A.; Ghoshal, S. Selective solubilization of polycyclic aromatic hydrocarbons from multicomponent nonaqueous-phase liquids into nonionic surfactant micelles. *Environ. Sci. Technol.* **2004**, *38*, 5878–5887. [\[CrossRef\]](#)
43. Yu, J.; Fischman, D.A.; Steck, T.L. Selective solubilization of proteins and phospholipids from red blood cell membranes by nonionic detergents. *J. Cell Biochem.* **1973**, *1*, 233–248. [\[CrossRef\]](#)
44. Gaudin, T.; Rotureau, P.; Pezron, I.; Fayet, G. New QSPR Models to Predict the Critical Micelle Concentration of Sugar-Based Surfactants. *Ind. Eng. Chem. Res.* **2016**, *55*, 11716–11726. [\[CrossRef\]](#)
45. Salager, J.-L.; Antón, R.; Anderez, J.M.; Aubry, J.-M. Formulation des micro-émulsions par la méthode HLD. *Techniques de l'Ingénieur* **2001**, *157*, 2001.
46. Acosta, E.; Szekeres, E.; Sabatini, D.A.; Harwell, J.H. Net-average curvature model for solubilization and supersolubilization in surfactant microemulsions. *Langmuir* **2003**, *19*, 186–195. [\[CrossRef\]](#)
47. Nguyen, T.T.; Edelen, A.; Neighbors, B.; Sabatini, D.A. Biocompatible lecithin-based microemulsions with rhamnolipid and sophorolipid biosurfactants: Formulation and potential applications. *J. Colloid Interface Sci.* **2010**, *348*, 498–504. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Rongsayamanont, W.; Soonglerdsongpha, S.; Khondee, N.; Pinyakong, O.; Tongcumpou, C.; Sabatini, D.A.; Luepromchai, E. Formulation of crude oil spill dispersants based on the HLD concept and using a lipopeptide biosurfactant. *J. Hazard. Mater.* **2017**, *334*, 168–177. [\[CrossRef\]](#)
49. Nguyen, T.T.; Morgan, C.; Poindexter, L.; Fernandez, J. Application of the Hydrophilic–Lipophilic Deviation Concept to Surfactant Characterization and Surfactant Selection for Enhanced Oil Recovery. *J. Surfactants Deterg.* **2019**. [\[CrossRef\]](#)
50. Sunwoo, C.K.; Wade, W.H. Optimal surfactant structures for cosurfactant-free microemulsion systems I. C16 and C14 Guerbet alcohol hydrophobes. *J. Dispers. Sci. Technol.* **1992**, *13*, 491–514. [\[CrossRef\]](#)
51. Zhang, G.; Yu, J.; Du, C.; Lee, R. Formulation of surfactants for very low/high salinity surfactant flooding without alkali. In Proceedings of the SPE International Symposium on Oilfield Chemistry, The Woodlands, TX, USA, 13–15 April 2015.
52. Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of different approaches to define the applicability domain of QSAR models. *Molecules* **2012**, *17*, 4791–4810. [\[CrossRef\]](#)
53. Wan, J.; Meng, D.; Long, T.; Ying, R.; Ye, M.; Zhang, S.; Li, Q.; Zhou, Y.; Lin, Y. Simultaneous Removal of Lindane, Lead and Cadmium from Soils by Rhamnolipids Combined with Citric Acid. *PLoS ONE* **2015**, *10*, e0129978. [\[CrossRef\]](#)
54. Zhong, H.; Liu, Y.; Liu, Z.; Jiang, Y.; Tan, F.; Zeng, G.; Yuan, X.; Yan, M.; Niu, Q.; Liang, Y. Degradation of pseudo-solubilized and mass hexadecane by a *Pseudomonas aeruginosa* with treatment of rhamnolipid biosurfactant. *Int. Biodeter. Biodegr.* **2014**, *94*, 152–159. [\[CrossRef\]](#)

55. Guo, Y.; Mulligan, C.N.; Nieh, M.P. An unusual morphological transformation of rhamnolipid aggregates induced by concentration and addition of styrene: A small angle neutron scattering (SANS) study. *Colloids Surf. A* **2011**, *373*, 42–50. [[CrossRef](#)]
56. Wan, J.; Chai, L.; Lu, X.; Lin, Y.; Zhang, S. Remediation of hexachlorobenzene contaminated soils by rhamnolipid enhanced soil washing coupled with activated carbon selective adsorption. *J. Hazard. Mater.* **2011**, *189*, 458–464. [[CrossRef](#)] [[PubMed](#)]
57. Shin, K.H.; Kim, J.Y.; Kim, K.W. Effect of Biosurfactant Addition on the Biodegradation of phenanthrene in soil-water system. *Environ. Eng. Res.* **2008**, *13*, 8–13. [[CrossRef](#)]
58. Bai, G.; Brusseau, M.L.; Miller, R.M. Influence of cation type, ionic strength, and pH on solubilization and mobilization of residual hydrocarbon by a biosurfactant. *J. Contam. Hydrol.* **1998**, *30*, 265–279. [[CrossRef](#)]
59. Zhang, Y.; Maier, W.J.; Miller, R.M. Effect of rhamnolipids on the dissolution, bioavailability, and biodegradation of phenanthrene. *Environ. Sci. Technol.* **1997**, *31*, 2211–2217. [[CrossRef](#)]
60. Li, S.; Pi, Y.; Bao, M.; Zhang, C.; Zhao, D.; Li, Y.; Sun, P.; Lu, J. Effect of rhamnolipid biosurfactant on solubilization of polycyclic aromatic hydrocarbons. *Mar. Pollut. Bull.* **2015**, *101*, 219–225. [[CrossRef](#)]
61. Zhang, X.; Guo, Q.; Hu, Y.; Lin, H. Effects of monorhamnolipid and dirhamnolipid on sorption and desorption of triclosan in sediment–water system. *Chemosphere* **2013**, *90*, 581–587. [[CrossRef](#)]
62. Zhu, S.; Liang, S.; You, X. Effect of Rhamnolipid Biosurfactant on Solubilization and Biodegradation of Polycyclic Aromatic Hydrocarbons. In Proceedings of the 2013 Third International Conference on Intelligent System Design and Engineering Applications, Hong Kong, China, 16–18 January 2013; pp. 651–654.
63. Yin, H.; Qiang, J.; Jia, Y.; Ye, J.; Peng, H.; Qin, H.; Zhang, N.; He, B. Characteristics of biosurfactant produced by *Pseudomonas aeruginosa* S6 isolated from oil-containing wastewater. *Process. Biochem.* **2009**, *44*, 302–308. [[CrossRef](#)]
64. Harendra, S.; Vipulanandan, C. Degradation of high concentrations of PCE solubilized in SDS and biosurfactant with Fe/Ni bi-metallic particles. *Colloids Surf. A* **2008**, *322*, 6–13. [[CrossRef](#)]
65. Bordas, F.; Lafrance, P.; Villemur, R. Conditions for effective removal of pyrene from an artificially contaminated soil using *Pseudomonas aeruginosa* 57SJ rhamnolipids. *Environ. Pollut.* **2005**, *138*, 69–76. [[CrossRef](#)]
66. Bodour, A.A.; Guerrero-Barajas, C.; Jiorle, B.V.; Malcomson, M.E.; Paull, A.K.; Somogyi, A.; Trinh, L.N.; Bates, R.B.; Maier, R.M. Structure and characterization of flavolipids, a novel class of biosurfactants produced by *Flavobacterium* sp. strain MTN11. *Appl. Environ. Microbiol.* **2004**, *70*, 114–120. [[CrossRef](#)] [[PubMed](#)]
67. Barkay, T.; Navon-Venezia, S.; Ron, E.; Rosenberg, E. Enhancement of solubilization and biodegradation of polyaromatic hydrocarbons by the bioemulsifier alasan. *Appl. Environ. Microbiol.* **1999**, *65*, 2697–2702. [[CrossRef](#)] [[PubMed](#)]
68. Thangamani, S.; Shreve, G.S. Effect of anionic biosurfactant on hexadecane partitioning in multiphase systems. *Environ. Sci. Technol.* **1994**, *28*, 1993–2000. [[CrossRef](#)] [[PubMed](#)]
69. OECD. *Guidance Document on the Validation of (Quantitative) Structure-Activity Relationships [(Q)SAR] Models*; OECD Environment Health and Safety Publications: Paris, France, 2007; Volume 69.
70. Aşçı, Y.; Nurbas, M.; Açikel, Y.S. Sorption of Cd (II) onto kaolin as a soil component and desorption of Cd (II) from kaolin using rhamnolipid biosurfactant. *J. Hazard. Mater.* **2007**, *139*, 50–56. [[CrossRef](#)]
71. Cohen, R.; Exerowa, D. Surface forces and properties of foam films from rhamnolipid biosurfactants. *Adv. Colloid Interface Sci.* **2007**, *134*, 24–34. [[CrossRef](#)]
72. Shin, K.-H.; Kim, K.-W.; Kim, J.-Y.; Lee, K.-E.; Han, S.-S. Rhamnolipid morphology and phenanthrene solubility at different pH values. *J. Environ. Qual.* **2008**, *37*, 509–514. [[CrossRef](#)]
73. Santos, S.F.; Zquette, D.; Fischer, H.; Itri, R. A systematic study of bovine serum albumin (BSA) and sodium dodecyl sulfate (SDS) interactions by surface tension and small angle X-ray scattering. *J. Colloid Interface Sci.* **2003**, *262*, 400–408. [[CrossRef](#)]
74. Edwards, A.L. *An Introduction to Linear Regression and Correlation*, 2nd ed.; W.H. Freeman: New York, NY, USA, 1984.
75. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics; Springer: New York, NY, USA, 2009.
76. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
77. Łozińska, N.; Głowacz-Różyńska, A.; Artichowicz, W.; Lu, Y.; Jungnickel, C. Microencapsulation of fish oil—determination of optimal wall material and encapsulation methodology. *J. Food Eng.* **2020**, *268*, 109730. [[CrossRef](#)]
78. Myers, D. *Surfaces, Interfaces, and Colloids: Principles and Applications*, 2nd ed.; VCH: New York, NY, USA, 1999.
79. Rangel-Yagui, C.O.; Pessoa, A., Jr.; Tavares, L.C. Micellar solubilization of drugs. *J. Pharm. Pharm. Sci.* **2005**, *8*, 147–163.
80. Brophy, J.; Bawden, D. Is Google enough? Comparison of an internet search engine with academic library resources. *Aslib Proc.* **2005**, *57*, 498–512. [[CrossRef](#)]
81. *Molinspiration Chemoinformatics*; Molinspiration: Bratislava, Slovak Republic, 2014.
82. *Addinsoft*, Version 2019.1.2; XLSTAT Statistical and Data Analysis Solution: Boston, MA, USA, 2019.
83. Jia, Q.; Zhao, Y.; Yan, F.; Wang, Q. QSAR model for predicting the toxicity of organic compounds to fathead minnow. *Environ. Sci. Pollut. Res.* **2018**, *25*, 35420–35428. [[CrossRef](#)] [[PubMed](#)]
84. Veerasamy, R.; Rajak, H.; Jain, A.; Sivadasan, S.; Varghese, C.P.; Agrawal, R.K. Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov.* **2011**, *3*, 511–519.
85. Linden, L.; Goss, K.-U.; Endo, S. 3D-QSAR predictions for  $\alpha$ -cyclodextrin binding constants using quantum mechanically based descriptors. *Chemosphere* **2017**, *169*, 693–699. [[CrossRef](#)] [[PubMed](#)]



86. Cherkassky, V.; Ma, Y. Selection of Meta-Parameters for Support Vector Regression. In Proceedings of the Artificial Neural Networks—ICANN 2002, Madrid, Spain, 28–30 August 2002. [\[CrossRef\]](#)
87. Cawley, G.C.; Talbot, N.L. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
88. Tobias, R.D. An Introduction to Partial Least Squares Regression. In *Proceedings of the Twentieth Annual SAS Users Group International Conference*; SAS Institute Inc: Cary, NC, USA, 1995; pp. 1250–1257.
89. Vinzi, V.E.; Chin, W.W.; Henseler, J.; Wang, H. *Handbook of Partial Least Squares*; Springer: New York, NY, USA, 2010; Volume 201.
90. Angrist, J.D.; Imbens, G.W.; Krueger, A.B. Jackknife instrumental variables estimation. *J. Appl. Econom.* **1999**, *14*, 57–67. [\[CrossRef\]](#)
91. Farrés, M.; Platikanov, S.; Tsakovski, S.; Tauler, R. Comparison of the variable importance in projection (VIP) and of the selectivity ratio (SR) methods for variable selection and interpretation. *J. Chemom.* **2015**, *29*, 528–536. [\[CrossRef\]](#)
92. Al-Bahry, S.; Al-Wahaibi, Y.; Elshafie, A.; Al-Bemani, A.; Joshi, S.; Al-Makhmari, H.; Al-Sulaimani, H. Biosurfactant production by *Bacillus subtilis* B20 using date molasses and its possible application in enhanced oil recovery. *Int. Biodeterior. Biodegr.* **2013**, *81*, 141–146. [\[CrossRef\]](#)
93. Luna, J.M.; Sarubbo, L.; Campos-Takaki, G.M. A new biosurfactant produced by *Candida glabrata* UCP 1002: Characteristics of stability and application in oil recovery. *Braz. Arch. Biol. Technol.* **2009**, *52*, 785–793. [\[CrossRef\]](#)
94. Thavasi, R.; Jayalakshmi, S.; Banat, I.M. Application of biosurfactant produced from peanut oil cake by *Lactobacillus delbrueckii* in biodegradation of crude oil. *Bioresour. Technol.* **2011**, *102*, 3366–3372. [\[CrossRef\]](#)
95. Luna, J.M.; Rufino, R.D.; Sarubbo, L.A.; Campos-Takaki, G.M. Characterisation, surface properties and biological activity of a biosurfactant produced from industrial waste by *Candida sphaerica* UCP0995 for application in the petroleum industry. *Colloids Surf. B* **2013**, *102*, 202–209. [\[CrossRef\]](#) [\[PubMed\]](#)

