

Article

Predicting Antifouling Activity and Acetylcholinesterase Inhibition of Marine-Derived Compounds Using a Computer-Aided Drug Design Approach

Susana P. Gaudêncio ^{1,2}  and Florbela Pereira ^{3,*} 

¹ Associate Laboratory i4HB—Institute for Health and Bioeconomy, NOVA School of Science and Technology, NOVA University of Lisbon, 2819-516 Caparica, Portugal; s.gaudencio@fct.unl.pt

² UCIBIO—Applied Molecular Biosciences Unit, Department of Chemistry, Blue Biotechnology and Biomedicine Lab, NOVA School of Science and Technology, NOVA University of Lisbon, 2819-516 Caparica, Portugal

³ LAQV, Department of Chemistry, NOVA School of Science and Technology, NOVA University of Lisbon, 2829-516 Caparica, Portugal

* Correspondence: florbela.pereira@fct.unl.pt

Abstract: Biofouling is the undesirable growth of micro- and macro-organisms on artificial water-immersed surfaces, which results in high costs for the prevention and maintenance of this process (billion €/year) for aquaculture, shipping and other industries that rely on coastal and off-shore infrastructure. To date, there are still no sustainable, economical and environmentally safe solutions to overcome this challenging phenomenon. A computer-aided drug design (CADD) approach comprising ligand- and structure-based methods was explored for predicting the antifouling activities of marine natural products (MNPs). In the CADD ligand-based method, 141 organic molecules extracted from the ChEMBL database and literature with antifouling screening data were used to build the quantitative structure–activity relationship (QSAR) classification model. An overall predictive accuracy score of up to 71% was achieved with the best QSAR model for external and internal validation using test and training sets. A virtual screening campaign of 14,492 MNPs from Encinar’s website and 14 MNPs that are currently in the clinical pipeline was also carried out using the best QSAR model developed. In the CADD structure-based approach, the 125 MNPs that were selected by the QSAR approach were used in molecular docking experiments against the acetylcholinesterase enzyme. Overall, 16 MNPs were proposed as the most promising marine drug-like leads as antifouling agents, e.g., macrocyclic lactam, macrocyclic alkaloids, indole and pyridine derivatives.

Keywords: marine natural products (MNPs); blue biotechnology; quantitative structure–activity relationship (QSAR); machine learning (ML) techniques; computer-aided drug design (CADD); molecular docking; virtual screening; antifouling activity; acetylcholinesterase enzyme (AChE)



Citation: Gaudêncio, S.P.; Pereira, F. Predicting Antifouling Activity and Acetylcholinesterase Inhibition of Marine-Derived Compounds Using a Computer-Aided Drug Design Approach. *Mar. Drugs* **2022**, *20*, 129. <https://doi.org/10.3390/md20020129>

Academic Editor: Orazio Tagliatalata-Scafati

Received: 13 December 2021

Accepted: 6 February 2022

Published: 8 February 2022

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



Copyright: © 2022 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/>).

1. Introduction

Marine biofouling is the undesired accumulation of micro-organisms, e.g., bacteria, cyanobacteria, unicellular algae and protozoa, and macro-organisms, e.g., seaweeds, barnacles, mussels and shells, on artificial water-immersed surfaces in a dynamic process that starts immediately after water submersion and can be a fast or slow process taking only hours or months to develop, respectively [1]. Marine biofouling creates risks to various industries, such as aquaculture and shipping, as well as for non-marine industries, e.g., paper manufacturing, food processing, underwater construction, power plants and others [2,3]. Settlement on the vessel’s hull results in damage to the rudder and propulsion systems [2,4], leads to an increasing drag of up to 60%, as well as a fuel consumption increase by 40%, increasing carbon dioxide and sulfur dioxide emissions [5] and the spread of nonindigenous

marine species into ecosystems worldwide, leading to environmental imbalances [6–10]. The most effective antifouling (AF) coatings contained biocides, such as tributyltin (TBT) and tributyltin oxide (TBTO), which were found to be harmful to non-target organisms and the environment [11] and thus were prohibited by the International Maritime Organization from Ship Surfaces in 2008, generating the demand for new generations of non-toxic or environment-friendly AF solutions [12–14].

Natural alternatives including primary or secondary metabolites isolated from marine organisms have been reported in several reviews to inhibit the settlement of different biofouling species [15–24]. The search for AF agents from marine sources began with bromo-derived metabolites, among the 2-furanone bromine derivatives extracted from red algae, which have been reported to prevent fouling [25], as well as bromopyrrole alkaloid derivatives with AF activity isolated from sponges (oroidin), inspiring the design of more than 50 synthetic analogues [26,27], and, more recently, antifouling bromotyrosine derivatives of the synoxazolidinone and the pulmonarin families [28]. Several studies reported MNPs with antifouling activity comprising the 2,5-diketopiperazine scaffold isolated from the marine sponge *Geodia barretti* [29], 6-benzyl and 6-isobutyl 2,5-diketopiperazine derivatives from marine-derived actinomycete *Streptomyces praecox* [30] and five diketopiperazines, cyclo-(L-Leu-L-Pro), cyclo-(L-Phe-L-Pro), cyclo-(L-Val-L-Pro), cyclo-(L-Trp-L-Pro) and cyclo-(L-Leu-L-Val), from deep-sea *Streptomyces fungicidicus* [31]. Comprising a meroterpenoid scaffold, napyradiomycin derivatives, isolated from marine-derived actinomycetes *Streptomyces aculeolatus*, were investigated by our group as antifouling inhibitors, having the advantage of inhibiting both micro- (antibiofilm activity) and macrofouling [32–34].

Computer-aided drug design (CADD) approaches have been used to guide decisions concerning the in vivo and in vitro testing of isolated NPs and extracts [35–39], to assist in the design of bioactive NP derivatives [40,41] and to virtually screen databases of known or proposed NPs [40,42–44]. To the best of our knowledge, the antifouling activity was quantitative structure–activity relationship (QSAR) modeled in only two previous works for the settlement of *Mytilus galloprovincialis* larvae [45,46]. Almeida et al. built two QSAR models using multilinear regression methods with, respectively, 19 and 16 nature-inspired (thio)xanthone [46] and chalcone [45] derivatives, including in vitro antifouling activity screening assays for the settlement of *Mytilus galloprovincialis* larvae.

Acetylcholinesterase (AChE) inhibitors are a class of drugs used for the treatment of Alzheimer's disease, glaucoma and autoimmune disorders [47–49]. The enzymes AChE [28] and tyrosinase (Tyr) were associated with the adhesive processes in the settlement of different biofouling species [28,46,50]. Almeida et al. reported a molecular docking study conducted by modulation of *Electrophorus electric* (fish) AChE of the two most promising (thio)xanthone antifouling agents [46]. Recently, Arabshahi et al. [50] reported an extensive virtual *Tetronarce californica* (fish) AChE homology screening campaign for 10,000 small organic molecules from the Chembridge library. The authors also reported the experimental screening of the most promising AChE inhibitors proposed by the in silico model, against five microfouling marine bacteria and marine microalgae macrofouling tunicate *Ciona savignyi*, discovering a potent novel inhibitor of tunicate settlement [50].

Herein, we report comprehensive computational modeling for the prediction of antifouling activities from two MNP libraries, by employing structure- and ligand-based CADD methodologies. The two libraries comprised 14,492 MNP from Prof. Encinar (<http://docking.umh.es/downloaddb>, accessed on 25 October 2021) and 14 MNPs from the clinical pipeline of MNPs (eight drugs approved and six MNPs in Phase II and III clinical trials). All the MNPs from the virtual screening libraries that were predicted to belong to the active class, i.e., 125 MNPs, were selected to proceed to the CADD structure-based method, where 125 MNPs selected by QSAR approach were screened by molecular docking against the AChE enzyme. In this CADD approach, a virtual screening hit list comprising 19 MNPs was assented based on some established thresholds, such as the probability of being active in the best antifouling model and the prediction of affinity between the AChE

of selected MNPs by molecular docking. A total of 16 MNPs have been proposed as the most promising marine drug-like leads as antifouling agents.

2. Results and Discussion

2.1. Chemical Space of the Antifouling Model

The whole data set (i.e., 141 small organic molecules) was randomly divided into a training set of 127 molecules (comprising 57 active and 70 inactive molecules) and a test set of 14 molecules (comprising six active and eight inactive molecules), which were used for the development and external validation of the QSAR classification models, respectively. The whole data set comprised seven structural classes or scaffold types, which are represented in Table 1 along with their antifouling activity classes and scaffold representative.

Table 1. Structural clusters and antifouling activity class counts within the seven structural clusters.

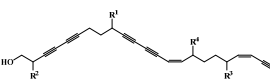
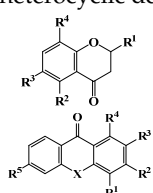
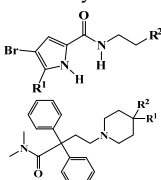
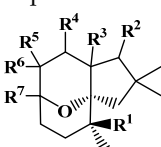
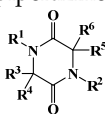
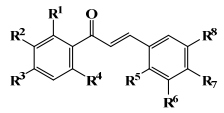
| Clusters ¹ | # ² (Active Class) | | Average MW (Da) ³ | | Average ALogP ⁴ | |
|----------------------------------------------------------------------------------------------------------------------|-------------------------------|-------|------------------------------|--------|----------------------------|------|
| | Tr | Te | Tr | Te | Tr | Te |
| I—acyclic derivative  | 11 (11) | 0 (0) | 361.65 | 0 | 2.86 | 0 |
| II—O-heterocyclic derivative  | 28 (9) | 3 (1) | 328.09 | 334.64 | 3.18 | 3.22 |
| III—N-heterocyclic derivative  | 19 (14) | 1 (0) | 363.92 | 493.04 | 2.50 | 3.65 |
| IV—terpenoid derivative  | 22 (5) | 6 (3) | 264.64 | 341.76 | 3.00 | 4.49 |
| V—diketopiperazine derivative  | 15 (10) | 3 (2) | 392.54 | 415.15 | 3.06 | 3.10 |

Table 1. Cont.

| Clusters ¹ | # ² (Active Class) | | Average MW (Da) ³ | | Average ALogP ⁴ | |
|-------------------------------------------------------------------------------------------------------------|-------------------------------|-------|------------------------------|--------|----------------------------|-------|
| | Tr | Te | Tr | Te | Tr | Te |
| VI—chalcone derivative  | 16 (3) | 0 (0) | 352.37 | 0 | 4.56 | 0 |
| VII—miscellaneous | 16 (5) | 1 (0) | 1164.53 | 975.69 | −0.88 | −1.57 |

¹ Cluster code and chemical structure of the cluster scaffold. ² Number of molecules in the training (Tr) and the test (Te) sets. ³ Molecular weight (MW) within the cluster for the training and test sets. ⁴ Octanol–water partition coefficient prediction within the cluster for the training and test sets.

All seven structural clusters (I, acyclic derivative, II, *O*-heterocyclic derivative, III, *N*-heterocyclic derivative, IV, terpenoid derivative, V, diketopiperazine derivative, VI, chalcone derivative, and VII, miscellaneous) were well represented in the training set, each comprising more than 10 molecules per class. The active class was more represented in three structural clusters with a percentage higher than 50%, namely I—acyclic derivative (100%), III—*N*-heterocyclic derivative (74%) and V—diketopiperazine derivative (67%). In the test set, only five structural clusters were represented, II–V and VII. In Table 1, the most representative scaffolds of the structural cluster are highlighted—for instance, for cluster I, a polyacetylene derivative; II, a chromone and a xanthone derivative; III, a pyrrole and a piperidine derivative; IV, a sesquiterpene derivative; V, a diketopiperazine, VI, a chalcone derivative; and VII, various scaffolds such as peptides and nature-inspired sulfated compounds. All clusters for the training and test sets, except for cluster VII, had an average MW value of less than 500 Da.

2.2. Establishment of QSAR Classification Model

Random Forests (RF) [51] were used to build models for antifouling prediction, exploring well-established PaDEL fingerprints (FPP and descriptors, e.g., five different types of FPs with different sizes (166 MACCS, MACCS keys; 307 Substructure; 881 PubChem fingerprints; 1024 CDK, circular fingerprints; 1024 CDK Ext, extended circular fingerprints with additional bits describing ring features) and 1376 1D&2D molecular descriptors (including electronic, topological and constitutional descriptors)) [52]. The performance of the models was successfully evaluated by internal validation (out-of-bag, OOB, estimation on the training set); see Table 2.

Table 2. Evaluation of the predictive performance of FPs and 1D&2D molecular descriptors for modeling the antifouling activity using the RF algorithm for the training set with an OOB estimation. The best models are highlighted in bold.

| Descriptors (#) | TP ¹ | TN ² | FN ³ | FP ⁴ | SE ⁵ | SP ⁶ | Q ⁷ | MCC ⁸ |
|----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|------------------|
| MACCS (166) ⁹ | 41 | 51 | 16 | 19 | 0.719 | 0.729 | 0.724 | 0.446 |
| Sub (307) ⁹ | 41 | 53 | 16 | 17 | 0.719 | 0.757 | 0.740 | 0.476 |
| PubChem (881) ⁹ | 43 | 48 | 14 | 22 | 0.754 | 0.686 | 0.717 | 0.438 |
| CDK (1024) ⁹ | 42 | 47 | 15 | 23 | 0.737 | 0.671 | 0.701 | 0.406 |
| ExtCDK (1024) ⁹ | 41 | 49 | 16 | 21 | 0.719 | 0.700 | 0.709 | 0.417 |
| 1D&2D (1376) | 40 | 53 | 17 | 17 | 0.702 | 0.757 | 0.732 | 0.459 |

¹ True positive. ² True negative. ³ False negative. ⁴ False positive. ⁵ Sensitivity, the ratio of true positive to the sum of true positive and false positive. ⁶ Specificity, the ratio of true negative to the sum of true negative and false negative. ⁷ Overall predictive accuracy, the ratio of the sum of true positive and true negative to the sum of true positive, true negative, false positive and false negative. ⁸ Matthews correlation coefficient. ⁹ Fingerprints, FPs.

From the seven sets of FPs and descriptors used to build the QSAR classification model, the best set for each type, fragment FPs (Sub), circular FPs (ExtCDK) and molec-

ular descriptors (1D&2D), were selected for further investigations; see Table 2. The 3D descriptors had a well-established relationship with biological activity and were expected to increase both the accuracy and robustness of the predictive models. After the exploration of models derived with molecular descriptors and FPs, we investigated the inclusion of 3D descriptors such as radial distribution function (RDF) descriptors (using a range of 128 and partial atomic charge as an atomic property) and the selection of descriptors using the RF descriptor importance parameter for the best three sets (Sub FPs, ExtCDK FPs and 1D&2D descriptors). Three sets of descriptors (Sub + RDF, ExtCDK + RDF and 1D&2D + RDF) as well as their selection were explored for modeling the antifouling activity using the RF algorithm in Table 3, where the results for the training set in OOB estimation are presented.

Table 3. Evaluation of the predictive performance of RDF descriptors and descriptor selection for modeling the antifouling activity using the RF algorithm for the training set with an OOB estimation. The best models are highlighted in bold.

| Model | # | SE ¹ | SP ² | Q ³ | MCC ⁴ |
|------------------------|------|-----------------|-----------------|----------------|------------------|
| Sub + RDF | 691 | 0.667 | 0.714 | 0.693 | 0.380 |
| Selection ⁵ | 50 | 0.667 | 0.714 | 0.693 | 0.380 |
| Selection ⁵ | 100 | 0.684 | 0.757 | 0.724 | 0.442 |
| Selection ⁵ | 150 | 0.702 | 0.786 | 0.748 | 0.489 |
| Selection ⁵ | 200 | 0.684 | 0.757 | 0.724 | 0.442 |
| ExtCDK + RDF | 1408 | 0.667 | 0.743 | 0.709 | 0.410 |
| Selection ⁵ | 12 | 0.754 | 0.729 | 0.740 | 0.481 |
| Selection ⁵ | 25 | 0.737 | 0.786 | 0.764 | 0.523 |
| Selection ⁵ | 50 | 0.702 | 0.771 | 0.740 | 0.474 |
| Selection ⁵ | 100 | 0.684 | 0.771 | 0.732 | 0.457 |
| 1D&2D + RDF | 1760 | 0.719 | 0.714 | 0.717 | 0.432 |
| Selection ⁵ | 50 | 0.807 | 0.800 | 0.803 | 0.605 |
| Selection ⁵ | 100 | 0.825 | 0.786 | 0.803 | 0.607 |
| Selection ⁵ | 150 | 0.807 | 0.800 | 0.803 | 0.605 |
| Selection ⁵ | 200 | 0.842 | 0.786 | 0.811 | 0.625 |
| Selection ⁵ | 250 | 0.772 | 0.800 | 0.787 | 0.571 |

¹ Sensitivity, the ratio of true positive to the sum of true positive and false positive. ² Specificity, the ratio of true negative to the sum of true negative and false negative. ³ Overall predictive accuracy, the ratio of the sum of true positive and true negative to the sum of true positive, true negative, false positive and false negative. ⁴ Matthews correlation coefficient. ⁵ The descriptor selection was evaluated based on the importance assigned by the RF model with the R program.

The 200 most important descriptors selected by the MeanDecreaseAccuracy parameter of the 1D&2D + RDF model were identified by the RF algorithm and enabled the training of a new RF model with better prediction accuracy in accordance with the Q and MCC values than the model trained with the whole set of descriptors (Table 3). A comparison of three machine learning (ML) techniques using the Weka software (support vector machines, SVM), R software (RF) and Keras software (deep learning multilayer perceptron networks, dMLP) for building the antifouling models with the 200 descriptors that were selected by the RF is shown in Table 4 for the test set.

Table 4. Exploration of different ML algorithms using the 200 selected descriptors.

| Model | SE ¹ | SP ² | Q ³ | MCC ⁴ |
|-------|-----------------|-----------------|----------------|------------------|
| RF | 0.667 | 0.750 | 0.714 | 0.417 |
| SVM | 0.830 | 0.500 | 0.643 | 0.344 |
| dMLP | 0.670 | 0.750 | 0.714 | 0.417 |

¹ Sensitivity, the ratio of true positive to the sum of true positive and false positive. ² Specificity, the ratio of true negative to the sum of true negative and false negative. ³ Overall predictive accuracy, the ratio of the sum of true positive and true negative to the sum of true positive, true negative, false positive and false negative. ⁴ Matthews correlation coefficient.

The best models were accomplished with the RF and α MLP algorithms using the 200 1D&2D + RDF selected descriptors, which achieved, for both models, a Q and MCC of 0.714 and 0.417 for the external test set. Majority voting predictions (consensus) were obtained by the RF, SVM and α MLP models (the consensus model, CM), and did not improve the results, with a Q and MCC of 0.571 and 0.167 for the test set; thus, in the next step of the virtual screening, we used the best model obtained, RF, with the 200 selected descriptors; see Tables 3 and 4).

The results obtained by the RF for the training and test sets that were in accordance with the seven structural clusters (I–VII), reported in Table 1, are shown in Table 5.

Table 5. The predictions of the best RF model by the seven structural clusters for the training and test sets. The best models are highlighted in bold.

| Cluster | # | SE ¹ | SP ² | Q ³ | MCC ⁴ |
|--------------|----|-----------------|-----------------|----------------|------------------|
| Training set | | | | | |
| I | 11 | 1.000 | - | 1.000 | 1.000 |
| II | 28 | 0.889 | 0.789 | 0.821 | 0.640 |
| III | 19 | 1.000 | 0.400 | 0.842 | 0.574 |
| IV | 22 | 0.800 | 0.941 | 0.909 | 0.741 |
| V | 15 | 0.900 | 0.000 | 0.600 | - |
| VI | 16 | 0.000 | 1.000 | 0.813 | - |
| VII | 16 | 0.400 | 0.812 | 0.688 | 0.234 |
| All | | 0.842 | 0.786 | 0.811 | 0.625 |
| Test set | | | | | |
| II | 3 | 1.000 | 1.000 | 1.000 | 1.000 |
| III | 1 | - | 1.000 | 1.000 | 1.000 |
| IV | 6 | 0.333 | 1.000 | 0.667 | 0.447 |
| V | 3 | 1.000 | 0.000 | 0.667 | - |
| VII | 1 | - | 0.000 | 0.000 | - |
| All | | 0.667 | 0.750 | 0.713 | 0.417 |

¹ Sensitivity, the ratio of true positive to the sum of true positive and false positive. ² Specificity, the ratio of true negative to the sum of true negative and false negative. ³ Overall predictive accuracy, the ratio of the sum of true positive and true negative to the sum of true positive, true negative, false positive and false negative. ⁴ Matthews correlation coefficient.

There were three structural clusters (I, II and IV, bold highlighted in Table 5) in which the predictions obtained were better than those obtained for the overall training set simultaneously considering the Q and MCC values. An improvement in the RF model prediction accuracies (Q = 0.821–1 and MCC = 0.64–1) was achieved for these three clusters of the training set, when compared with the prediction accuracy obtained for all the molecules of the training set (Q = 0.811 and MCC = 0.625). For the clusters II and V–VII, lower prediction accuracies were obtained, Q = 0.6–0.842 and MCC = 0.234–0.574. Interestingly, the best achieved predictions for structural clusters I and II were related to the best performance obtained for the active class prediction, with SE values of 1 and 0.889, respectively, compared to the SE value of 0.842 for all training sets. For example, for the test set, the average of the Prob_active (a_Prob_active) obtained by the active molecules predicted by the model as active, i.e., true positive (TP), was 0.59, which compares with the value of a_Prob_active of 0.54 obtained by the predicted molecules by the model as false positives (FP). The same relationship was obtained for molecules predicted as true negatives (FN) and false negatives (FN), with an a_Prob_active of 0.44 and 0.48, respectively. Additionally, it appears that, with a Prob_active higher than 0.59, there was no error in the prediction and all molecules predicted as active were active.

2.3. Analysis of Fingerprints and Descriptors Identified as Relevant for Modeling the Antifouling Activity

The selected 200 descriptors included 164 1D&2D (115 topological descriptors, 48 count type descriptors and one constitutional descriptor (Mannhold LogP, logarithm of the octanol–water partition coefficient)) and 36 RDF 3D descriptors (12 of type a (a positive

and a negative charge), 12 of type b (two positive charges) and 12 of type c (two negative charges)). The 1D&2D descriptors comprised 72 autocorrelation topological descriptors, which were 50 Broto–Moreau, 12 Moran and 10 Geary autocorrelation descriptors, weighted by mass, charges, van der Waals volumes, Sanderson electronegativities, polarizabilities, first ionization potential or I-state. Other topological descriptors, such as 6 Barysz matrices, 24 Burden-modified eigenvalues, 1 Detour matrix, 2 MDEs, 2 path counts, 3 topological charges, 3 distance matrices, 1 walk count descriptor and 1 weighted path descriptor, were also presented. The count type descriptors included 28 electrotopological state atom types, 10 extended topochemical atoms and 10 information content descriptors. A comparison of the best twenty 1D&2D + RDF molecular descriptors selected by descriptor importance of RF was used to build the QSAR classification models, which are presented in Tables 3 and 4, and these were analyzed and are presented in Figure 1.

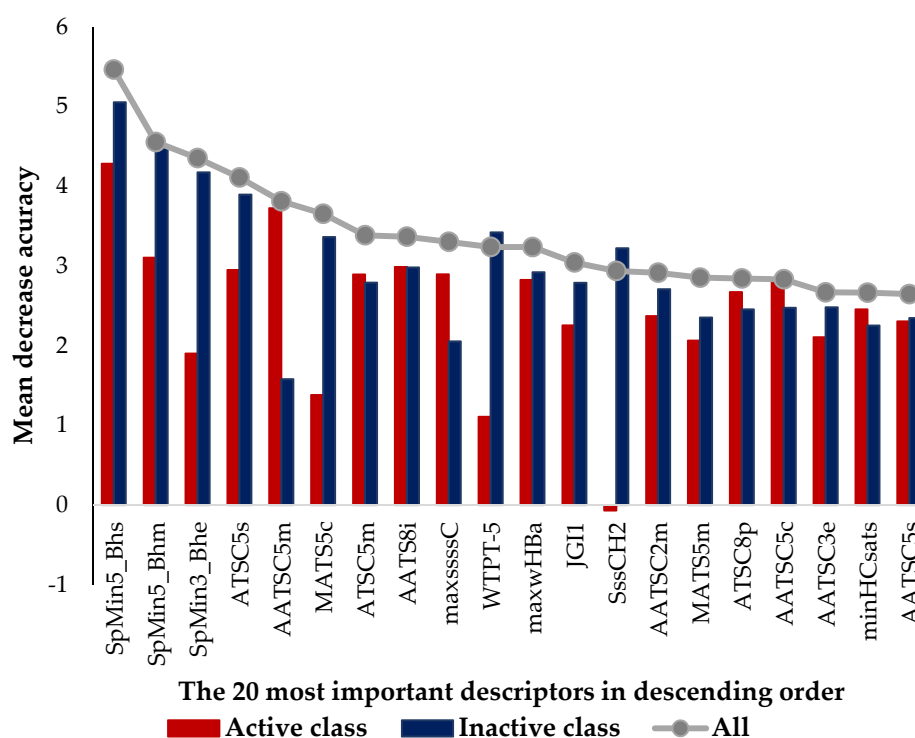


Figure 1. The twenty most important 1D&2D +RDF descriptors selected in RF classification models, where the first three descriptors in terms of importance are three Burden-modified eigenvalue descriptors weighted by relative I-state, mass and Sanderson electronegativities, respectively; there are several Broto–Moreau autocorrelations 4th–5th, 7th–8th, 14th, 16th–18th, 20th weighted by I-state, mass, mass, first ionization potential, mass, polarizabilities, charge, Sanderson electronegativities and I-state; two Moran autocorrelation descriptors, 6th and 15th weighted by charge and mass, respectively; four electrotopological state atom type descriptors, 9th (>C<), 11th (weak hydrogen bond acceptors), 13th (-CH2-), 19th (H bonded to B, Si, P, Ge, As, Se, Sn or P); one PaDEL weighted path descriptor, 10th (sum of path lengths starting from nitrogens); and one topological charge descriptor, 12th (mean topological charge index of order 1).

Interestingly, no 3D RDF descriptor appeared in the list of the twenty most important descriptors and the first RDF descriptor appeared only in the 30th position (two positive charges). Moreover, there were only seven out of the twenty most important descriptors that were more relevant in discriminating the active class, namely AATSC5m (5th), ATSC5m (7th), AATS8i (8th), maxssssC (9th), ATSC8p (16th), AATSC5c (17th) and minHCsats (19th). Of the nine Broto–Moreau autocorrelation descriptors existing in the list of the top 20, five of them were more relevant to discriminate the active class and, on the other hand, they also presented a lag higher than or equal to 5, which was related to a greater distance between

the structural features of interest. In contrast, the four Broto–Moreau autocorrelation descriptors that were more relevant for the inactive class presented a lag lower than or equal to 5. The three most important descriptors in the top 20 list were three Burden-modified eigenvalue descriptors and all of them were most relevant in the inactive class discrimination. This eigenvalue was suggested as an index of molecular branching, the smallest values corresponding to chain graphs (SpMin3_Bhe) and the highest to the most branched graphs (SpMin5_Bhs and SpMin5_Bhm) [53]. A very interesting behavior was observed with the two electrotopological state atom types, maxssssC (maximum atom-type E-state: >C<) and SssCH2 (sum of atom-type E-state: -CH₂-), which were more relevant for the active and inactive classes, respectively. The maxssssC descriptor encodes the maximum number of quaternary or asymmetric carbon atoms and could be seen as encoding structural complexity. On the other hand, the SssCH2 descriptor encoded the saturation of the molecule. Another very important descriptor to discriminate mainly the inactive class is the PaDEL weighted path descriptor, WTPT-5, which is the sum of all path weights starting from nitrogen atoms, revealing nitrogen-specific branching information. In agreement with the present work, the two QSAR studies reported by Almeida et al. highlighted the descriptors related to the branching, complexity and the influence of the molecule's interatomic distance for the modeling of the antifouling activity [45,46].

2.4. Application of the *In Silico* Antifouling QSAR Model in Virtual Screening

A virtual screening campaign was carried out to search for new lead-like antifouling inhibitors. The best QSAR model, the RF model, was selected for the virtual screening procedure using 14,492 MNPs from Prof. Encinar's website and 14 MNPs in the pharmaceutical pipeline (eight approved drugs and six MNPs in Phase II and III of clinical trials). The antifouling virtual screening of the MNP library in the pharmaceutical pipeline allowed us to assess the possibility of repurposing drugs of marine origin. Of these 14 MNPs from the pharmaceutical pipeline, only one MNP in Phase II of clinical trials presented activity against AChE, GTS-21 (DMXBA), a derivative of the NP, 2,4-dimethoxybenzylidene anabaseine dihydrochloride. There were 13,902 MNPs that were predicted to be active by the best QSAR model, of which 8349 MNPs were predicted to be active with a Prob_active greater than 0.59 (limit defined for the test set for which there are no prediction errors). From these MNPs, 5 (one approved drug and four MNPs in Phase II and III of clinical trials) and 8344 MNPs were from the pharmaceutical pipeline and from Encinar's database, respectively. Interestingly, of the five MNPs from the MNP pharmaceutical pipeline predicted to be active with the highest Prob_active was DMXBA with a value of 0.658. A more demanding limit has been defined for the CADD structure-based approach: all the MNPs from the virtual screening libraries that were predicted as belonging to the active class with a Prob_active greater than or equal to 0.68 were selected for molecular docking experiments. In the CADD structure-based method, the 125 MNPs selected by the QSAR classification approach were screened by molecular docking against acetylcholinesterase enzyme (AChE).

The list of eleven lead-like AChE inhibitors against antifouling activity generated from the AChE homology virtual screening, which were experimentally screened in *in vitro* and micro- and macrofouling assays reported by Arabshahi et al. [50], was used in this study as a second virtual screening library (Supplementary Data, Table S5). Only one out of the eleven lead-like AChE inhibitors was predicted to have antifouling activity with a Prob_active higher than 0.59 (Table S5), the morpholine derivative (Figure 2), in which experimental antifouling activity IC₅₀ = 16 µg/mL was reported (51.7 µM) [50]. However, none of the eleven compounds passed the established threshold, which was more demanding (Prob_active ≥ 0.68), to be selected for the molecular docking experiments.

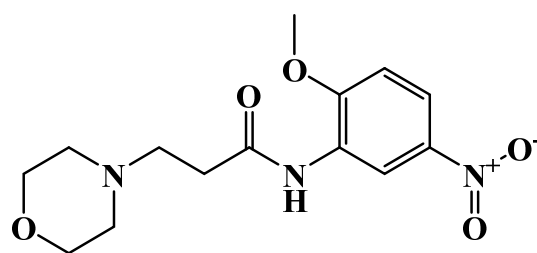


Figure 2. Chemical structure of the morpholine derivative.

2.5. Molecular Docking against AChE Enzyme

The 125 MNPs from Encinar's database selected by the QSAR classification approach were screened by molecular docking against AChE enzyme (PDB ID: 6TT0) [54]. The antifouling agents, synoxazolidinone A, synoxazolidinone C and donepezil, known as AChE inhibitors [28], were used as positive controls and the phenolic derivative that was predicted to not have antifouling activity in virtual screening was used as a negative control in the molecular docking experiments. A list of virtual screening hits comprising 19 MNPs was approved based on molecular docking experiments, in which a threshold of $\Delta G_B \leq -7$ kcal/mol was established for predicting the affinity between AChE and selected MNPs. To prioritize the best marine drug-like leads as antifouling AChE inhibitors from the list of 19 selected MNPs by the antifouling QSAR model and molecular docking of AChE enzyme, the absorption, distribution, metabolism, excretion and toxicity (ADMET) properties were predicted via in silico methods using the pKCSM software (<http://biosig.unimelb.edu.au/pkcsml/>, accessed on 25 October 2021) [55]. Sixteen MNPs, a macrocyclic lactam (CAS 156310-18-8), seven macrocyclic alkaloids (CAS 126622-63-7, 126622-64-8, 156310-18-8, 155944-26-6, 157536-35-1, 105305-54-2 and 105418-77-7), seven indole derivatives (CAS 142677-10-9, 134029-43-9, 134029-44-0, 134029-45-1, 142677-09-6, 223596-72-3, 134779-34-3) and a pyridine derivative (CAS 59697-14-2) were proposed as marine drug-like leads as antifouling AChE inhibitors. Three MNPs were excluded due to their predicted toxicity to fish, namely against flathead minnows. The Autodock Vina software (<http://vina.scripps.edu/>, accessed on 25 October 2021) [56] was used to perform the flexible virtual screening of the 125 MNPs to find the most favorable binding interactions, and the calculated free binding energies by the set of search space coordinates are reported in Table 6 for the 16 MNPs selected, and the positive (synoxazolidinone A and C; donepezil, an AChE inhibitor used for Alzheimer disease therapy) and the negative (phenolic derivative derivative) controls.

Table 6. Structures and calculated free binding energies (ΔG_B , in kcal/mol) of the sixteen selected MNPs, the positive (synoxazolidinone A and C; donepezil) and negative (phenolic derivative) controls.

| CAS | Chemical Structure | Name/Structural Category | Natural Source | Prob_A | ΔG_B (kcal/mol) ¹ |
|-------------|--------------------|------------------------------------|----------------------------|--------|--------------------------------------|
| 147362-39-8 | | cylindramide/lactam | marine sponge ² | 0.684 | -11.3 |
| 126622-63-7 | | haliclamine B/macrocyclic alkaloid | marine sponge ³ | 0.682 | -8.2 |

Table 6. Cont.

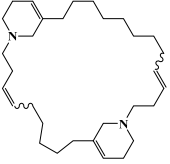
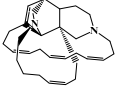
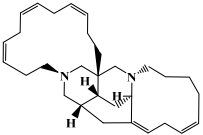
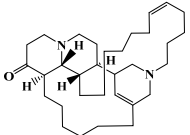
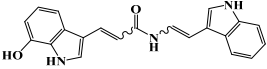
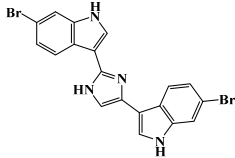
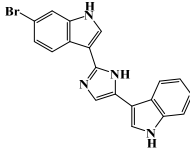
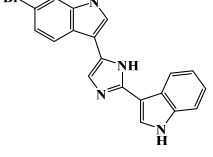
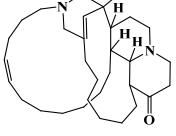
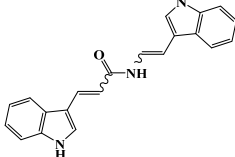
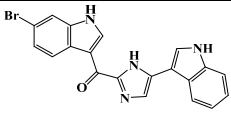
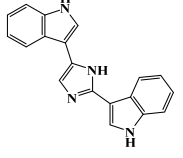

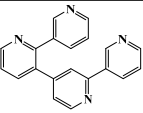
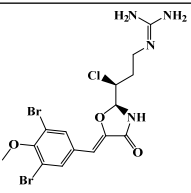
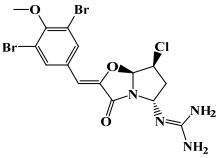
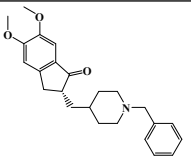
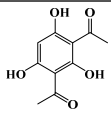
| CAS | Chemical Structure | Name/Structural Category | Natural Source | Prob_A | ΔG_B (kcal/mol) ¹ |
|-------------|-------------------------------------------------------------------------------------|------------------------------------|----------------------------|--------|--------------------------------------|
| 126622-64-8 |  | haliclamine A/macrocylic alkaloid | marine sponge ³ | 0.682 | −7.8 |
| 156310-18-8 |  | ingamine B/macrocylic alkaloid | marine sponge ⁴ | 0.682 | −7.8 |
| 155944-26-6 |  | madangamines A/macrocylic alkaloid | marine sponge ⁴ | 0.694 | −7.7 |
| 105305-54-2 |  | serain 3/macrocylic alkaloid | marine sponge ⁵ | 0.686 | −7.5 |
| 142677-10-9 |  | chondriamide B/indole | red alga ⁶ | 0.682 | −7.5 |
| 134029-43-9 |  | nortopsentin A/indole | marine sponge ⁷ | 0.702 | −7.3 |
| 134029-44-0 |  | nortopsentin B/indole | marine sponge ⁷ | 0.698 | −7.3 |
| 134029-45-1 |  | nortopsentin C/indole | marine sponge ⁷ | 0.700 | −7.3 |
| 105418-77-7 |  | serain 1/macrocylic alkaloid | marine sponge ⁵ | 0.686 | −7.2 |
| 142677-09-6 |  | chondriamide A/indole | red alga ⁶ | 0.682 | −7.2 |

Table 6. Cont.

| CAS | Chemical Structure | Name/Structural Category | Natural Source | Prob_A | ΔG_B (kcal/mol) ¹ |
|------------------|-------------------------------------------------------------------------------------|------------------------------------------|----------------------------|--------|--------------------------------------|
| 223596-72-3 |  | isobromodeoxytopsentin/ indole | marine sponge ⁸ | 0.680 | −7.2 |
| 134779-34-3 |  | nortopsentin D/indole | marine sponge ⁷ | 0.688 | −7.1 |
| 157536-35-1 |  | keramaphidin B/macrocylic alkaloid | marine sponge ⁹ | 0.684 | −7.1 |
| 59697-14-2 |  | nemertelline/ pyridine | marine worm ¹⁰ | 0.680 | −7.0 |
| positive control |  | synoxazolidinone A | - | - | −6.5 |
| positive control |  | synoxazolidinone C | - | - | −6.7 |
| positive control |  | donepezil | - | - | −6.5 |
| negative control |  | phenolic | - | - | −5.1 |

¹ AChE enzyme: center X: 25.435 Y: 69.621 Z: 278.986; ² *Halichondria cylindrata*; ³ *Haliclona* sp.; ⁴ *Xestospongia ingens*; ⁵ *Reniera sarai*; ⁶ *Chondria* sp.; ⁷ *Spongisorites ruetzleri* and *Haliclona* sp.; ⁸ *Spongisorites* sp.; ⁹ *Amphimedon* sp.; ¹⁰ *Amphiporus angulatus*.

The prediction of the ADMET properties of the sixteen selected MNPs by the antifouling QSAR model and molecular docking of AChE enzyme is presented in Table S1, in the Supplementary Materials. In Figure 3, the interaction profile of the best-docked pose for the two most probable lead-like antifouling AChE inhibitors, a lactam derivative—cylindramide—and a macrocyclic alkaloid—haliclamine B—is represented.

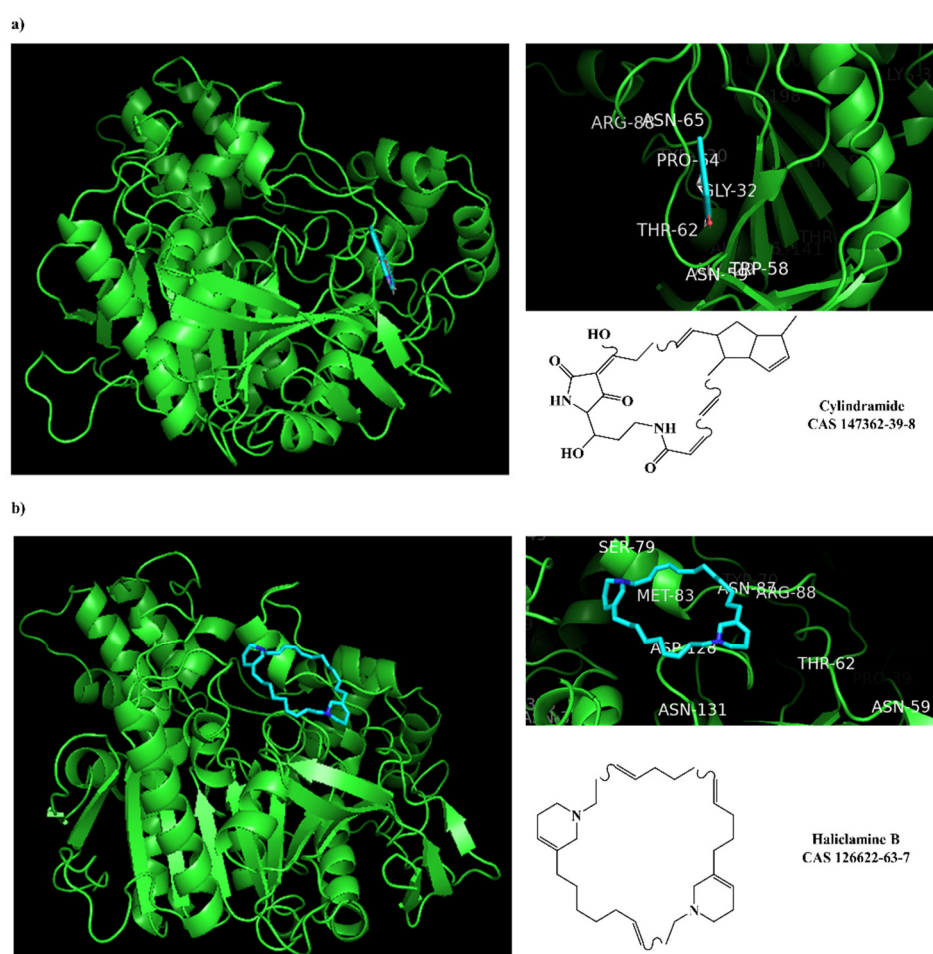


Figure 3. Interaction profiles of the best-docked poses for the two hits (a) cylindramide and (b) haliclamine B.

New scoring functions based on more precise physics-based descriptors to better represent the protein–ligand recognition process have been developed. DockThor, a web service for molecular docking simulation (<https://dockthor.lncc.br/v2/>, accessed on 6 January 2022), was used to perform molecular docking of the two best macrocycle hits (cylindramine and haliclamine B), the best non-macrocycle hit (indole derivative, CAS 142677-10-9) and the positive and negative controls against the AChE enzyme (PDB ID: 6TT0). In DockThor, a set of new empirical scoring functions to estimate protein–ligand binding affinity were developed by explicitly accounting for physics-based interaction terms based on the MMFF94S force field combined with ML [57]. The DockThor scores obtained for the two best macrocycle hits (cylindramine and haliclamine B), the best non-macrocycle hit (indole derivative, CAS 142677-10-9) and the positive (synoxazolidinone A and C; donepezil) and negative (phenolic derivative) controls were -8.508 kcal/mol (-11.3 kcal/mol using Autodock Vina), -7.008 kcal/mol (-8.2 kcal/mol using Autodock Vina), -8.634 kcal/mol (-7.5 kcal/mol using Autodock Vina), -7.749 kcal/mol (-6.5 kcal/mol using Autodock Vina), -7.56 kcal/mol (-6.7 kcal/mol using Autodock Vina) and -6.416 kcal/mol (-5.1 kcal/mol using Autodock Vina), respectively. The interaction profiles of the best-docked poses predicted by DockThor for the two best macrocycle hits, the best non-macrocycle hit and the positive and negative controls are presented in Figure 4.

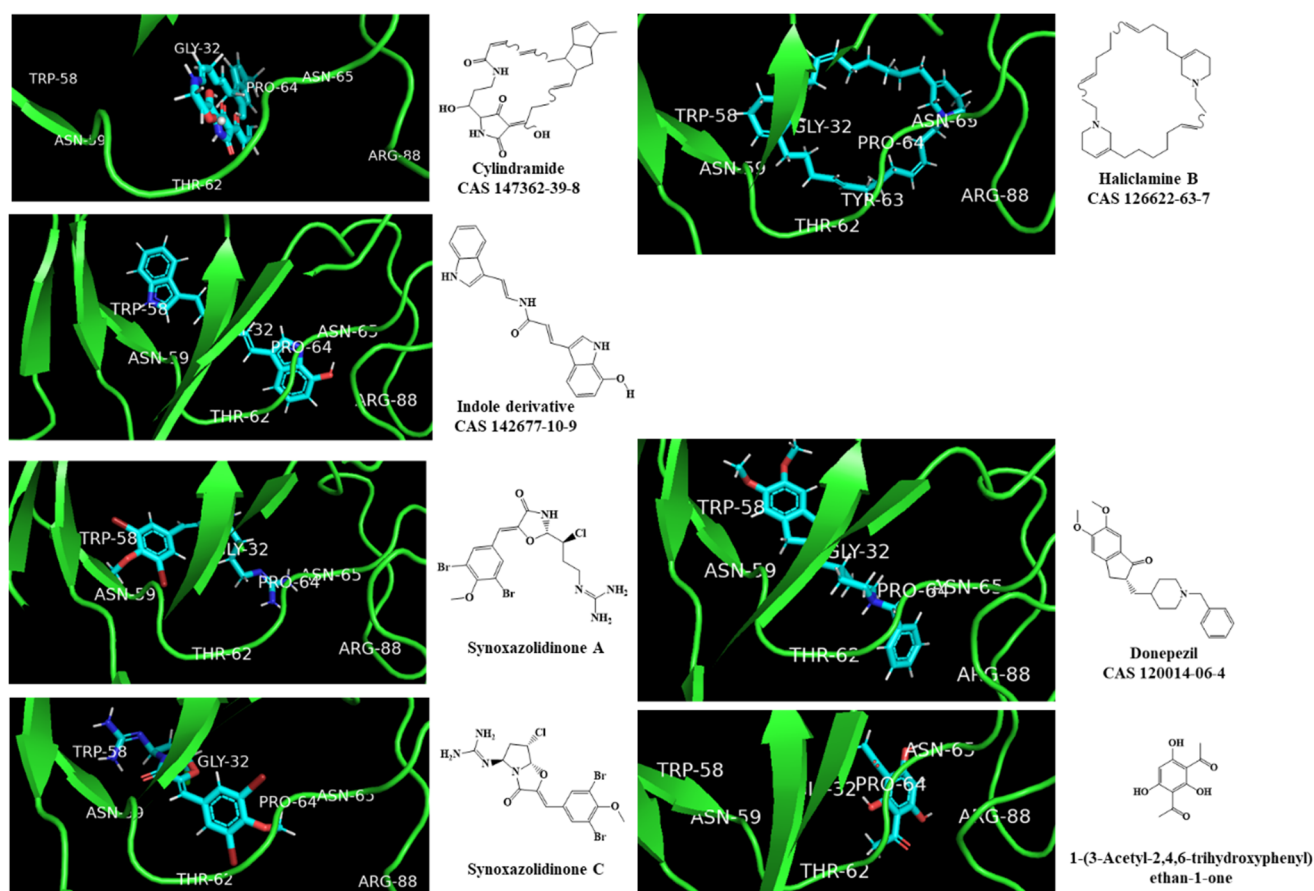


Figure 4. Interaction profiles of the best-docked poses for the two macrocyclic hits (cylindramide and haliclamine B), the best non-macrocycle hit (indole derivative) and the positive (synoxazolidinone A and C; donepezil) and negative (phenolic derivative) controls.

The peripheral anionic site (PAS) of AChE is composed of five residues (TYR-70, ASP-72, TYR-121, TRP-279 and TYR-334) and is involved in the allosteric modulation of catalysis at the active center [46]. This site is the target of various anti-cholinesterase inhibitors. In this work, other residues (e.g., ARG-88, ASN-65, PRO-64, GLY-32, THR-62, TRP-58 and ASN-59) forming the hydrophobic interactions in the PAS pocket are highlighted in Figures 3 and 4. The binding of donepezil to the PAS of AChE is in accordance with its proposed peculiar inhibitory mechanism, which involves a reversible double-binding site interaction with the catalytic anionic site and PAS of the enzyme [54]. Unlike our approach and in other reported studies [46,54], Arabshahi et al. [50] performed a virtual screening by molecular docking of AChE at the catalytic anionic site and not at the PAS. Although none of the 11 reported compounds [50] passed the QSAR model threshold to be subjected to molecular docking, we still performed the molecular docking and the docking scores are presented in Table S5 (Supplementary Data). It was verified that none of these compounds exceeded the established threshold in the molecular docking experiments, $\Delta\text{GB} \leq -7$ kcal/mol.

3. Materials and Methods

3.1. Data Sets/Selection of Training and Test Sets

The antifouling data set comprising 142 molecules, 63 and 79 organic molecules, was extracted from the ChEMBL (<https://www.ebi.ac.uk/chembl/>, accessed on 21 July 2021) [58] and by searching in the literature indexed in the Web of Science Core Collection until June 2021, respectively. The ChEMBL data set was obtained by searching for marine organisms with antifouling activity, such as barnacles (e.g., *Balanus amphitrite*), mussels (e.g., *Mytilus galloprovincialis*), bushy bryozoan (e.g., *Bugula neritina*) and marine algae (e.g., *Ulva conglobata*). The

antifouling activity was classified using two activity classes: (A, active)—inhibition % > 52% and EC₅₀, IC₅₀ ≤ 25 µg/mL; (B, inactive)—inhibition % ≤ 52% and EC₅₀, IC₅₀ > 25 µg/mL. After collecting these data sets, the duplicates were removed based on the IUPAC international chemical identifier (InChI) codes; however, the chirality was considered, and racemic compounds (or cases where no stereochemistry was indicated) were considered as one of the possible stereoisomers. Thereafter, the final data set comprised 141 organic molecules and was divided into a training set comprising 127 molecules (class A, 57 molecules and class B, 70 molecules) and a test set comprising 14 molecules (class A, 6 molecules and class B, 8 molecules). The partitioning of the data set into training and testing sets was performed randomly according to the composition of the antifouling classes (active and inactive). The composition of the 10 structural categories shown in Table 1 was not considered. The built QSAR models were developed and externally validated using the training and test sets, respectively.

The virtual data set comprised 14,492 MNPs from Prof. Encinar's website (<http://docking.umh.es/downloaddb>, accessed on 25 October 2021) saved in the MDL SDF data format and 14 MNPs from the pharmaceutical pipeline set (eight approved drugs and six MNPs in Phase II and III of clinical trials). Three duplicates with the training and test sets were removed and the final virtual data set comprised 14,503 molecules.

A second virtual library comprising eleven lead-like AChE inhibitors against antifouling activity reported by Arabshahi et al. [50] was also used.

SMILES strings of the data sets, and the corresponding experimental and predicted activities, are available as Supplementary Data, Tables S2, S3 and S5.

3.2. Calculation of Descriptors

The molecular structures of molecules in all data sets were standardized by normalizing tautomeric and mesomeric groups and by removing small disconnected fragments using the JChem Standardizer tool, version 5.7.13.0 (ChemAxon Ltd., Budapest, Hungary). The optimization of the three-dimensional molecular structures was carried out with CORINA version 2.4 (Molecular Networks GmbH, Erlangen, Germany). PaDEL-Descriptor (Pharmaceutical Data Exploration Laboratory, Singapore) version 2.21 (<http://www.yapcwsoft.com/dd/padeldescriptor/>, accessed on 21 July 2021) [52] was used to calculate empirical molecular fingerprints (FPs) and 1D&2D molecular descriptors. FPs of various types were calculated and exploited to build QSAR models, namely 166 MACCS (MACCS keys), 307 Substructure (presence and count of SMARTS patterns for Laggner functional group classification—Sub), 881 PubChem fingerprints (ftp://ftp.ncbi.nlm.nih.gov/pubchem/specifications/pubchem_fingerprints.txt, accessed on 21 July 2021), 1024 CDK (circular fingerprints) and 1024 CDK extended (Ext circular fingerprints with additional bits describing ring features). The 1D&2D molecular descriptors comprised descriptors of various types, including electronic, topological and constitutional descriptors, in a total of 1376 descriptors. Radial distribution function (RDF) pair descriptors [59] and 3D RDF descriptors were calculated by sampling the function of Equation (1) at 128 equally distributed values of r between 0 and 12.8 Å:

$$RDF(r) = \sum_{i=1}^{N-1} \sum_{j=i+1}^N p_i p_j e^{-B(r-r_{ij})^2} \quad (1)$$

where N is the number of atoms in the molecule, p_i is the charge of atom i , B is a fuzziness parameter (it was 100 in this study), and r_{ij} is the 3D distance between atoms i and j . The *RDF* descriptors were separated into three sets of 128 descriptors per pair of atoms with (a) one positive and one negative charge, (b) two positive charges and (c) two negative charges. The partial atomic charges—natural bond orbital (NBO) partial atomic charges—were estimated using an ML tool developed by Aires-de-Sousa and co-workers (<http://joao.airesdesousa.com/charges>, accessed on 21 July 2021) [60].

3.3. Selection of Descriptors and Optimization of QSAR Models

In the quest for QSAR models with the minimum possible number of descriptors, descriptor selection was performed based on the importance of descriptors assessed by the RF (computeAttributeImportance) algorithm [51] implemented in the R program [61]. Selection of descriptors was accomplished using this procedure, with the importance of descriptors assessed by RF within an OOB methodology using the 12, 25, 50, 100, 150, 200 and 250 most important descriptors and RF algorithm as an ML technique employing the following statistical metrics: true positives (TP), true negatives (TN), false positives (FP), false negatives (FN), sensitivity (SE, prediction accuracy for active antifouling molecules), specificity (SP, prediction accuracy for inactive antifouling molecules), overall predictive accuracy (Q) and Matthews correlation coefficient (MCC).

3.4. Class Balancer

In general, class imbalance is more demanding for ML algorithms and this imbalance introduces a bias due to their preference for the majority class [62]. Our antifouling activity training set was unbalanced, and the imbalance ratio was 1:1.22 for the A: active and B: inactive classes, respectively. To solve this problem, the classes were balanced using the RF *samplesize* parameter with R version 3.6.1. [61]. This parameter was set to be of the same size as the minority class (active class). With this parameter, some molecules belonging to the minority class were used more than once.

3.5. Machine Learning (ML) Methods

3.5.1. Random Forest (RF)

The RF model [51,63] was built from a set of unpruned classification trees, which were created using bootstrap samples from the training set. For each individual tree, the best split at each node was defined using a randomly selected subset of descriptors. Each of the individual classification trees was created using different training and validation sets. The final prediction of the model resulted from the majority vote of classification trees in the forest. Model performance was evaluated internally with the prediction error for molecules left out in the bootstrap procedure (OOB estimation). The method quantifies the importance of a descriptor by the increase in misclassification that occurs when descriptor values are randomly permuted, correlated with the mean decrease in the precision parameter. RFs also assigned a probability to every prediction based on the number of votes obtained by the predicted class. RFs were grown with the R program [61], version 3.6.1, using the random forest library [64]. As a result of the nature of the two-class imbalance, this problem was alleviated by defining the class weight ranges of 1–57 and 1–57 for classes A and B, respectively, using the *samplesize* parameter.

3.5.2. Support Vector Machines (SVMs)

SVMs [65] map the training data into a hyperspace through a nonlinear mapping (a boundary or hyperplane) and then separate the classes of objects in this space. The examples of the training set—the support vectors—allowed us to position the boundary. To transform data into a hyperspace where classes become linearly separable, kernel functions were used. In this study, SVMs were implemented with Scikit-learn [66] using the LIBSVM package [67]. The type of SVM was set to C-SVM-classification and the radial basis function was used for the kernel function. Hyperparameter tuning was performed using ten-fold cross-validation with the GridSearchCV tool. C and γ values varied in the range of 1×10^{-2} to 1×10^{13} and 1×10^{-9} to 1×10^{13} , respectively. In total, 10,000 experiments were performed. The C and γ values were finally set to 1×10^7 and 1×10^{-8} , respectively, and the other parameters were used with default values. To alleviate the imbalanced two-class problem, the *class_weight* parameter was set to be “balanced”, in which the smaller class was replicated until it had as many molecules as in the larger one class.

3.5.3. Deep Learning Multilayer Perceptron Networks (dMLP)

The feed-forward neural networks were implemented using the open-source software library Keras [68] version 2.2.5 based on the Tensorflow numerical backend engine [69]. These popular software tools, written in Python, make it easy to develop and apply deep neural networks; however, the main challenge in applying dMLP is the design of an adequate network architecture. After several experiments, the final optimal hyperparameter settings were selected for our study based on 10-fold cross-validation experiments with the training set and are listed in Table 7.

Table 7. Hyperparameter settings of the best dMLP model.

| Hyperparameter | Setting |
|---------------------------------------------|---------------------|
| Initializer | Glorot uniform |
| Number of hidden layers | 2 |
| Number of neurons in the 1st and 2nd layers | 200 |
| Number of neurons in the 3rd | 2 |
| Activation 1st–2nd layers | Relu |
| Activation 3rd layer | Sigmoid |
| Batch size | 36 |
| Optimizer | Adadelta |
| Loss | Binary crossentropy |
| Epochs | 100 |

3.6. Molecular Docking

The virtual screening using the best QSAR model, the RF classification model using the 200 most important 1D&2D + RDF molecular descriptors, allowed the prioritization of a list of the 125 MNP virtual screening hits. OpenBabel software (version 2.3.1, freely available under an open-source license from <http://openbabel.org>, accessed on 21 July 2021) [70] was used to convert mol2 files into PDBQT files. PDBQT files were used for coupling to the AChE enzyme with Autodock Vina (version 1.1, Center for Computational Structural Biology, Scripps Research Institute, CA, USA) [56]. The macromolecule coupling target was the AChE enzyme from *Tetronarce californica* (PDB ID: 6TT0) [54]. Water molecules, carbohydrate molecules and ligands (1R, 3S-cis- and 1S, 3R-cis-donepezil derived enantiomers) were removed from 6TT0 [54] prior to docking using AutoDockTools (<http://mglttools.scripps.edu/>, accessed on 21 July 2021). During enzyme preparation, 6TT0, explicit hydrogen atoms and Gasteiger charges for each atom were added. Autodock Vina performed a flexible molecular docking in which the target's conformation was considered a rigid unit while the ligands were flexible and adaptable to the target. Autodock Vina looked for the lowest binding affinity conformations and returned ten different conformations for each ligand. The search space coordinates of the AChE enzyme were maximized to allow the entire macromolecule to be considered for docking. The search space coordinates were center X: 25.179 Y: 72.212 Z: 281.175; dimensions X: 20,000 Y: 20,000 Z: 20,000. AChE enzyme ligand tethering was performed by regulating the parameters of the genetic algorithm (GA), using 10 runs of the GA criteria. DockThor, a web service for molecular docking simulation (<https://dockthor.lncc.br/v2/>, accessed on 6 January 2022), was used to perform molecular docking of the two best macrocycle hits (cylindramine and haliclamine B), the best non-macrocycle hit (indole derivative, CAS 142677-10-9) and the positive and negative controls against AChE enzyme (PDB ID: 6TT0) [57]. The search space coordinates were center X: 25.179 Y: 72.212 Z: 281.175; dimensions X: 20,000 Y: 20,000 Z: 20,000. AChE enzyme ligand tethering was performed by regulating the parameters of the GA, using 12,750 and 500,000 runs, population size and number of evaluations of the GA criteria, respectively.

The docking binding poses were visualized with PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC). Docking scores of 125 virtual hits against the AChE enzyme are shown in Table S4, Supplementary Data.

4. Conclusions

A CADD approach relying on ligand- and structure-based methodologies was successfully used to predict new inhibitory MNPs against antifouling AChE. Two MNPs, cylindramide (CAS 147362-39-8) and haliclamine B (CAS 126622-63-7), were proposed as the most promising marine drug-like leads as antifouling AChE inhibitors. To the best of our knowledge, the CADD ligand-based study using a QSAR classification model, developed here in this study, is the largest study ever performed with regard both to the number of molecules involved and to the number of structural families involved in the modeling of the antifouling activity, and the best model achieved an overall predictive accuracy score of up to 71% for both test and training sets. In future works, the proposed sixteen marine drug-like leads against antifouling AChE enzyme may be validated experimentally. These results enabled us to build virtual libraries of marine-derived drug-like leads, which may be virtually screened using the best antifouling QSAR model and molecular docking against the AChE enzyme. In addition, for MNPs that are experimentally confirmed to have antifouling activity, the AChE inhibitory mechanism will be studied to determine the type of action, e.g., reversible interaction with both the catalytic anionic site and the PAS, sterically blocking ligands from entering and leaving the active site gorge and allosteric alteration of the catalytic triad conformation.

Supplementary Materials: The following data are available online at <https://www.mdpi.com/article/10.3390/md20020129/s1>, Tables S1–S5 (XLSX). The following files are available free of charge. SMILES strings of the data set (training and test sets), the corresponding experimental and predicted activities are available as Supplementary Materials, Tables S2 and S3, respectively. Moreover, SMILES strings of the 14,492 MNPs from Encinar's website and MNPs clinical pipeline sets, for the virtual screening data set, the corresponding predicted activities are available as Supplementary Materials, Table S4. Predictions of ADMET properties with in silico methods, using the pKCSM software for a list of 16 selected MNPs by QSAR antifouling model and molecular docking of AChE enzyme are available as Supplementary Materials, Table S1. The list of eleven lead-like AChE inhibitors by Arabshahi et al. [50], the corresponding experimental, predicted activities and docking scores against the AChE enzyme are available as Supplementary Materials, Table S5.

Author Contributions: Conceptualization: F.P. and S.P.G.; Methodology: F.P.; Software: F.P.; Validation: F.P. (in silico modeling), S.P.G. (pharmaceutical pipeline data); Formal Analysis: F.P. (in silico modeling); Investigation: F.P. (in silico modeling) and S.P.G. (pharmaceutical pipeline data and MNP meroterpenoid library); Resources: F.P. (in silico modeling) and S.P.G. (pharmaceutical pipeline data and MNP meroterpenoid library); Writing—Original Draft Preparation: F.P. and S.P.G.; Writing—Review and Editing: F.P. and S.P.G.; Funding Acquisition: F.P. and S.P.G. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support from Fundação para a Ciência e Tecnologia (FCT) Portugal, under grant UIDB/50006/2020 (provided to the Associate Laboratory for Green Chemistry LAQV), is greatly appreciated. F.P. thanks Fundação para a Ciência e a Tecnologia, MCTES, for the Norma transitória DL 57/2016 Program Contract. This work is financed by national funds from FCT—Fundação para a Ciência e a Tecnologia, I.P., in the scope of the project UIDP/04378/2020 of the Research Unit on Applied Molecular Biosciences—UCIBIO and the project LA/P/0140/2020 of the Associate Laboratory Institute for Health and Bioeconomy—i4HB.

Data Availability Statement: Data are contained within the article or Supplementary Material.

Acknowledgments: We thank ChemAxon Ltd. for access to JChem and Marvin.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Magin, C.M.; Cooper, S.P.; Brennan, A.B. Non-toxic antifouling strategies. *Mater. Today* **2010**, *13*, 36–44. [CrossRef]
2. Schultz, M.P.; Bendick, J.A.; Holm, E.R.; Hertel, W.M. Economic impact of biofouling on a naval surface ship. *Biofouling* **2011**, *27*, 87–98. [CrossRef] [PubMed]
3. Schultz, M.P. Effects of coating roughness and biofouling on ship resistance and powering. *Biofouling* **2007**, *23*, 331–341. [CrossRef]

4. Schultz, M.P.; Walker, J.M.; Steppe, C.N.; Flack, K.A. Impact of diatomaceous biofilms on the frictional drag of fouling-release coatings. *Biofouling* **2015**, *31*, 759–773. [[CrossRef](#)] [[PubMed](#)]
5. Bhushan, B. Biomimetics: Lessons from nature—An overview. *Philos. Trans. Royal Soc. A* **2009**, *367*, 1445–1486. [[CrossRef](#)] [[PubMed](#)]
6. Ware, C.; Berge, J.; Sundet, J.H.; Kirkpatrick, J.B.; Coutts, A.D.M.; Jelmert, A.; Olsen, S.M.; Floerl, O.; Wisz, M.S.; Alsos, I.G. Climate change, non-indigenous species and shipping: Assessing the risk of species introduction to a high-Arctic archipelago. *Divers. Distrib.* **2014**, *20*, 10–19. [[CrossRef](#)]
7. Ashton, G.V.; Davidson, I.C.; Geller, J.; Ruiz, G.M. Disentangling the biogeography of ship biofouling: Barnacles in the Northeast Pacific. *Glob. Ecol. Biogeogr.* **2016**, *25*, 739–750. [[CrossRef](#)]
8. Pettengill, J.B.; Wendt, D.E.; Schug, M.D.; Hadfield, M.G. Biofouling likely serves as a major mode of dispersal for the polychaete tubeworm *Hydroides elegans* as inferred from microsatellite loci. *Biofouling* **2007**, *23*, 161–169. [[CrossRef](#)]
9. Piola, R.F.; Johnston, E.L. The potential for translocation of marine species via small-scale disruptions to antifouling surfaces. *Biofouling* **2008**, *24*, 145–155. [[CrossRef](#)]
10. Yamaguchi, T.; Prabowo, R.E.; Ohshiro, Y.; Shimono, T.; Jones, D.; Kawai, H.; Otani, M.; Oshino, A.; Inagawa, S.; Akaya, T.; et al. The introduction to Japan of the Titan barnacle, *Megabalanus coccopoma* (Darwin, 1854) (Cirripedia: Balanomorpha) and the role of shipping in its translocation. *Biofouling* **2009**, *25*, 325–333. [[CrossRef](#)] [[PubMed](#)]
11. Sonak, S.; Pangam, P.; Giriyani, A.; Hawaldar, K. Implications of the ban on organotins for protection of global coastal and marine ecology. *J. Environ. Manag.* **2009**, *90*, S96–S108. [[CrossRef](#)]
12. Callow, J.A.; Callow, M.E. Trends in the development of environmentally friendly fouling-resistant marine coatings. *Nat. Commun.* **2011**, *2*, 244. [[CrossRef](#)] [[PubMed](#)]
13. Kirschner, C.M.; Brennan, A.B. Bio-Inspired Antifouling Strategies. *Annu. Rev. Mater. Res.* **2012**, *42*, 211–229. [[CrossRef](#)]
14. Chambers, L.D.; Stokes, K.R.; Walsh, F.C.; Wood, R.J.K. Modern approaches to marine antifouling coatings. *Surf. Coat. Technol.* **2006**, *201*, 3642–3652. [[CrossRef](#)]
15. Othmani, A.; Bunet, R.; Bonnefont, J.L.; Briand, J.F.; Culioli, G. Settlement inhibition of marine biofilm bacteria and barnacle larvae by compounds isolated from the Mediterranean brown alga *Taonia atomaria*. *J. Appl. Phycol.* **2016**, *28*, 1975–1986. [[CrossRef](#)]
16. Sathesh, S.; Ba-akdah, M.A.; Al-Sofyani, A.A. Natural antifouling compound production by microbes associated with marine macroorganisms—A review. *Electron. J. Biotechnol.* **2016**, *21*, 26–35. [[CrossRef](#)]
17. Almeida, J.R.; Vasconcelos, V. Natural antifouling compounds: Effectiveness in preventing invertebrate settlement and adhesion. *Biotechnol. Adv.* **2015**, *33*, 343–357. [[CrossRef](#)] [[PubMed](#)]
18. Qian, P.-Y.; Li, Z.; Xu, Y.; Li, Y.; Fusetani, N. Mini-review: Marine natural products and their synthetic analogs as antifouling compounds: 2009–2014. *Biofouling* **2015**, *31*, 101–122. [[CrossRef](#)] [[PubMed](#)]
19. Qian, P.-Y.; Xu, Y.; Fusetani, N. Natural products as antifouling compounds: Recent progress and future perspectives. *Biofouling* **2010**, *26*, 223–234. [[CrossRef](#)]
20. Dobretsov, S.; Dahms, H.U.; Qian, P.Y. Inhibition of biofouling by marine microorganisms and their metabolites. *Biofouling* **2006**, *22*, 43–54. [[CrossRef](#)]
21. Wang, K.-L.; Wu, Z.-H.; Wang, Y.; Wang, C.-Y.; Xu, Y. Mini-Review: Antifouling Natural Products from Marine Microorganisms and Their Synthetic Analogs. *Mar. Drugs* **2017**, *15*, 266. [[CrossRef](#)]
22. Qi, S.-H.; Ma, X. Antifouling Compounds from Marine Invertebrates. *Mar. Drugs* **2017**, *15*, 263. [[CrossRef](#)] [[PubMed](#)]
23. Dahms, H.U.; Dobretsov, S. Antifouling Compounds from Marine Macroalgae. *Mar. Drugs* **2017**, *15*, 265. [[CrossRef](#)] [[PubMed](#)]
24. Moodie, L.W.K.; Sepcic, K.; Turk, T.; Frangez, R.; Svenson, J. Natural cholinesterase inhibitors from marine organisms. *Nat. Prod. Rep.* **2019**, *36*, 1053–1092. [[CrossRef](#)]
25. Dworjanyn, S.A.; de Nys, R.; Steinberg, P.D. Chemically mediated antifouling in the red alga *Delisea pulchra*. *Mar. Ecol. Prog. Ser.* **2006**, *318*, 153–163. [[CrossRef](#)]
26. Richards, J.J.; Ballard, T.E.; Huigens, R.W., III; Melander, C. Synthesis and screening of an oroidin library against *Pseudomonas aeruginosa* biofilms. *Chembiochem* **2008**, *9*, 1267–1279. [[CrossRef](#)]
27. Melander, C.; Moeller, P.D.R.; Ballard, T.E.; Richards, J.J.; Huigens, R.W., III; Cavanagh, J. Evaluation of dihydrooroidin as an antifouling additive in marine paint. *Int. Biodeterior. Biodegradation* **2009**, *63*, 529–532. [[CrossRef](#)] [[PubMed](#)]
28. Trepos, R.; Cervin, G.; Hellio, C.; Pavia, H.; Stensen, W.; Stensvag, K.; Svendsen, J.-S.; Haug, T.; Svenson, J. Antifouling Compounds from the Sub-Arctic Ascidian *Synoicum pulmonaria*: Synoxazolidinones A and C, Pulmonarins A and B, and Synthetic Analogues. *J. Nat. Prod.* **2014**, *77*, 2105–2113. [[CrossRef](#)] [[PubMed](#)]
29. Sjogren, M.; Goransson, U.; Johnson, A.L.; Dahlstrom, M.; Andersson, R.; Bergman, J.; Jonsson, P.R.; Bohlin, L. Antifouling activity of brominated cyclopeptides from the marine sponge *Geodia barretti*. *J. Nat. Prod.* **2004**, *67*, 368–372. [[CrossRef](#)] [[PubMed](#)]
30. Cho, J.Y.; Kang, J.Y.; Hong, Y.K.; Baek, H.H.; Shin, H.W.; Kim, M.S. Isolation and Structural Determination of the Antifouling Diketopiperazines from Marine-Derived *Streptomyces praecox* 291-11. *Biosci. Biotechnol. Biochem.* **2012**, *76*, 1116–1121. [[CrossRef](#)]
31. Li, X.; Dobretsov, S.; Xu, Y.; Xiao, X.; Hung, O.S.; Qian, P.-Y. Antifouling diketopiperazines produced by a deep-sea bacterium, *Streptomyces fungicidicus*. *Biofouling* **2006**, *22*, 201–208. [[CrossRef](#)] [[PubMed](#)]
32. Prieto-Davo, A.; Dias, T.; Gomes, S.E.; Rodrigues, S.; Parera-Valadez, Y.; Borralho, P.M.; Pereira, F.; Rodrigues, C.M.P.; Santos-Sanches, I.; Gaudencio, S.P. The Madeira Archipelago As a Significant Source of Marine-Derived Actinomycete Diversity with Anticancer and Antimicrobial Potential. *Front. Microbiol.* **2016**, *7*, 1594. [[CrossRef](#)] [[PubMed](#)]

33. Bauermeister, A.; Pereira, F.; Grilo, I.R.; Godinho, C.C.; Paulino, M.; Almeida, V.; Gobbo-Neto, L.; Prieto-Davo, A.; Sobral, R.G.; Lopes, N.P.; et al. Intra-clade metabolomic profiling of MAR4 *Streptomyces* from the Macaronesia Atlantic region reveals a source of anti-biofilm metabolites. *Environ. Microbiol.* **2019**, *21*, 1099–1112. [[CrossRef](#)] [[PubMed](#)]
34. Pereira, F.; Almeida, J.R.; Paulino, M.; Grilo, I.R.; Macedo, H.; Cunha, I.; Sobral, R.G.; Vasconcelos, V.; Gaudencio, S.P. Antifouling Napyradiomycins from Marine-Derived Actinomycetes *Streptomyces aculeolatus*. *Mar. Drugs* **2020**, *18*, 63. [[CrossRef](#)] [[PubMed](#)]
35. Cruz, S.; Gomes, S.E.; Borrallho, P.M.; Rodrigues, C.M.P.; Gaudencio, S.P.; Pereira, F. In Silico HCT116 Human Colon Cancer Cell-Based Models En Route to the Discovery of Lead-Like Anticancer Drugs. *Biomolecules* **2018**, *8*, 56. [[CrossRef](#)]
36. Dias, T.; Gaudencio, S.P.; Pereira, F. A Computer-Driven Approach to Discover Natural Product Leads for Methicillin-Resistant *Staphylococcus aureus* Infection Therapy. *Mar. Drugs* **2019**, *17*, 16. [[CrossRef](#)] [[PubMed](#)]
37. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Don Duy, N.; Watrous, J.; Kapon, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [[CrossRef](#)]
38. Lang, G.; Mayhudin, N.A.; Mitova, M.I.; Sun, L.; van der Sar, S.; Blunt, J.W.; Cole, A.L.J.; Ellis, G.; Laatsch, H.; Munro, M.H.G. Evolving trends in the dereplication of natural product extracts: New methodology for rapid, small-scale investigation of natural product extracts. *J. Nat. Prod.* **2008**, *71*, 1595–1599. [[CrossRef](#)]
39. Camp, D.; Davis, R.A.; Campitelli, M.; Ebdon, J.; Quinn, R.J. Drug-like Properties: Guiding Principles for the Design of Natural Product Libraries. *J. Nat. Prod.* **2012**, *75*, 72–81. [[CrossRef](#)]
40. Gaudencio, S.P.; Pereira, F. A Computer-Aided Drug Design Approach to Predict Marine Drug-Like Leads for SARS-CoV-2 Main Protease Inhibition. *Mar. Drugs* **2020**, *18*, 633. [[CrossRef](#)]
41. Wang, L.; Le, X.; Li, L.; Ju, Y.C.; Lin, Z.X.; Gu, Q.; Xu, J. Discovering New Agents Active against Methicillin-Resistant *Staphylococcus aureus* with Ligand-Based Approaches. *J. Chem. Inf. Model.* **2014**, *54*, 3186–3197. [[CrossRef](#)] [[PubMed](#)]
42. Pereira, F.; Aires-de-Sousa, J. Computational Methodologies in the Exploration of Marine Natural Product Leads. *Mar. Drugs* **2018**, *16*, 236. [[CrossRef](#)] [[PubMed](#)]
43. Pereira, F. Have marine natural product drug discovery efforts been productive and how can we improve their efficiency? *Expert Opin. Drug Discov.* **2019**, *14*, 717–722. [[CrossRef](#)] [[PubMed](#)]
44. Llanos, M.A.; Gantner, M.E.; Rodriguez, S.; Alberca, L.N.; Bellera, C.L.; Talevi, A.; Gavernet, L. Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: The Main Protease (Mpro) Case Study. *J. Chem. Inf. Model.* **2021**, *61*, 3758–3770. [[CrossRef](#)]
45. Almeida, J.R.; Moreira, J.; Pereira, D.; Pereira, S.; Antunes, J.; Palmeira, A.; Vasconcelos, V.; Pinto, M.; Correia-da-Silva, M.; Cidade, H. Potential of synthetic chalcone derivatives to prevent marine biofouling. *Sci. Total Environ.* **2018**, *643*, 98–106. [[CrossRef](#)] [[PubMed](#)]
46. Almeida, J.R.; Palmeira, A.; Campos, A.; Cunha, I.; Freitas, M.; Felpeto, A.B.; Turkina, M.V.; Vasconcelos, V.; Pinto, M.; Correia-da-Silva, M.; et al. Structure-Antifouling Activity Relationship and Molecular Targets of Bio-Inspired(thio)xanthenes. *Biomolecules* **2020**, *10*, 1126. [[CrossRef](#)] [[PubMed](#)]
47. Tadesse, M.; Svenson, J.; Sepicic, K.; Trembleau, L.; Engqvist, M.; Andersen, J.H.; Jaspars, M.; Stensvag, K.; Haug, T. Isolation and Synthesis of Pulmonarins A and B, Acetylcholinesterase Inhibitors from the Colonial Ascidian *Synoicum pulmonaria*. *J. Nat. Prod.* **2014**, *77*, 364–369. [[CrossRef](#)]
48. Kaur, J.; Zhang, M.Q. Molecular modelling and QSAR of reversible acetylcholinesterase inhibitors. *Curr. Med. Chem.* **2000**, *7*, 273–294. [[CrossRef](#)]
49. Munoz-Torrero, D. Acetylcholinesterase Inhibitors as Disease-Modifying Therapies for Alzheimer’s Disease. *Curr. Med. Chem.* **2008**, *15*, 2433–2455. [[CrossRef](#)]
50. Arabshahi, H.J.; Trobec, T.; Foulon, V.; Hellio, C.; Frangez, R.; Sepicic, K.; Cahill, P.; Svenson, J. Using Virtual AChE Homology Screening to Identify Small Molecules With the Ability to Inhibit Marine Biofouling. *Front. Mar. Sci.* **2021**, *8*, 762287. [[CrossRef](#)]
51. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
52. Yap, C.W. PaDEL-Descriptor: An Open Source Software to Calculate Molecular Descriptors and Fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [[CrossRef](#)]
53. Todeschini, R.; Consonni, V. *Molecular Descriptors for Chemoinformatics*; WILEY-VCH: Weinheim, Germany, 2009; Volumes 1–2.
54. Catto, M.; Pisani, L.; de la Mora, E.; Belviso, B.D.; Mangiatordi, G.F.; Pinto, A.; De Palma, A.; Denora, N.; Caliandro, R.; Colletier, J.-P.; et al. Chiral Separation, X-ray Structure, and Biological Evaluation of a Potent and Reversible Dual Binding Site AChE Inhibitor. *ACS Med. Chem. Lett.* **2020**, *11*, 869–876. [[CrossRef](#)]
55. Pires, D.E.V.; Blundell, T.L.; Ascher, D.B. pkCSM: Predicting Small-Molecule Pharmacokinetic and Toxicity Properties Using Graph-Based Signatures. *J. Med. Chem.* **2015**, *58*, 4066–4072. [[CrossRef](#)] [[PubMed](#)]
56. Trott, O.; Olson, A.J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [[CrossRef](#)]
57. Guedes, I.A.; Barreto, A.M.S.; Marinho, D.; Krempser, E.; Kuenemann, M.A.; Sperandio, O.; Dardenne, L.E.; Miteva, M.A. New machine learning and physics-based scoring functions for drug discovery. *Sci. Rep.* **2021**, *11*, 1–19. [[CrossRef](#)]
58. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magarinos, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards direct deposition of bioassay data. *Nucleic Acids Res.* **2019**, *47*, D930–D940. [[CrossRef](#)]
59. Selzer, P.; Ertl, P. Identification and classification of GPCR ligands using self-organizing neural networks. *QSAR Comb. Sci.* **2005**, *24*, 270–276. [[CrossRef](#)]

60. Zhang, Q.; Zheng, F.; Fartaria, R.; Latino, D.A.R.S.; Qu, X.; Campos, T.; Zhao, T.; Aires-de-Sousa, J. A QSPR approach for the fast estimation of DFT/NBO partial atomic charges. *Chemom. Intell. Lab. Syst.* **2014**, *134*, 158–163. [CrossRef]
61. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing: Vienna, Austria. 2014. Available online: <http://www.R-project.org> (accessed on 21 July 2021).
62. Jain, S.; Kotsampasakou, E.; Ecker, G.F. Comparing the performance of meta-classifiers—a case study on selected imbalanced data sets relevant for prediction of liver toxicity. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 583–590. [CrossRef]
63. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inform. Comput. Sci.* **2003**, *43*, 1947–1958. [CrossRef] [PubMed]
64. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22.
65. Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
66. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
67. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*. [CrossRef]
68. Chollet, F.K. GitHub, Seattle, WA, USA. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 21 July 2021).
69. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv* **2016**, arXiv:1603.04467.
70. O’Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33. [CrossRef]