

Article

# PTML Modeling for Pancreatic Cancer Research: In Silico Design of Simultaneous Multi-Protein and Multi-Cell Inhibitors

Valeria V. Kleandrova<sup>1</sup>  and Alejandro Speck-Planche<sup>2,\*</sup> 

<sup>1</sup> Laboratory of Fundamental and Applied Research of Quality and Technology of Food Production, Moscow State University of Food Production, Volokolamskoe Shosse 11, 125080 Moscow, Russia; valeria.kleandrova@gmail.com

<sup>2</sup> Grupo de Química Computacional y Teórica (QCT-USFQ), Departamento de Ingeniería Química, Universidad San Francisco de Quito, Diego de Robles y vía Interoceánica, Quito 170901, Ecuador

\* Correspondence: alejspivanovich@gmail.com

**Abstract:** Pancreatic cancer (PANC) is a dangerous type of cancer that is a major cause of mortality worldwide and exhibits a remarkably poor prognosis. To date, discovering anti-PANC agents remains a very complex and expensive process. Computational approaches can accelerate the search for anti-PANC agents. We report for the first time two models that combined perturbation theory with machine learning via a multilayer perceptron network (PTML-MLP) to perform the virtual design and prediction of molecules that can simultaneously inhibit multiple PANC cell lines and PANC-related proteins, such as caspase-1, tumor necrosis factor-alpha (TNF-alpha), and the insulin-like growth factor 1 receptor (IGF1R). Both PTML-MLP models exhibited accuracies higher than 78%. Using the interpretation from one of the PTML-MLP models as a guideline, we extracted different molecular fragments desirable for the inhibition of the PANC cell lines and the aforementioned PANC-related proteins and then assembled some of those fragments to form three new molecules. The two PTML-MLP models predicted the designed molecules as potentially versatile anti-PANC agents through inhibition of the three PANC-related proteins and multiple PANC cell lines. Conclusions: This work opens new horizons for the application of the PTML modeling methodology to anticancer research.

**Keywords:** caspase-1; cell line; fragment; IGF1R; MLP; multi-target; pancreatic cancer; TNF-alpha; virtual design



**Citation:** Kleandrova, V.V.; Speck-Planche, A. PTML Modeling for Pancreatic Cancer Research: In Silico Design of Simultaneous Multi-Protein and Multi-Cell Inhibitors. *Biomedicines* **2022**, *10*, 491. <https://doi.org/10.3390/biomedicines10020491>

Academic Editor: Satoshi Wada

Received: 29 January 2022

Accepted: 15 February 2022

Published: 18 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

Pancreatic cancer (PANC) is currently recognized as the seventh most significant cause of cancer-related deaths worldwide. In addition to being associated with a very poor prognosis, PANC presents five highly alarming aspects. First, the mortality rate of PANC almost equals its incidence rate, since PANC accounted for 458,918 new cases and 432,242 deaths in 2018 [1]. This is consistent with the overall 5-year survival rate of approximately 6%, which makes PANC the most lethal cancer of all [2]. Second, the global burden involving this intractable neoplasm has more than doubled over the past 25 years [3]. Third, PANC is characterized by the emergence of drug resistance [4]; this makes PANC difficult to treat. Four, from a genetic point of view, PANC is very complex because PANC predisposition genes are not well understood, although several genes suggested to be involved in PANC development and progression have been studied [5]. Last, the current chemotherapeutic drugs used to treat PANC are limited in terms of effectiveness because they are mainly available as adjuvants and act through specific mechanisms of action. All these aspects indicate the urgent need for new and more effective anti-PANC chemotherapeutics able to act as multi-target drugs, simultaneously inhibiting several PANC-related proteins.

To date, some of the biomolecular targets studied in PANC research are proteins such as caspase-1, tumor necrosis factor-alpha (TNF-alpha), and insulin-like growth factor-1

receptor (IGF1R), which are promising biomolecular targets against PANC. In the case of caspase-1 and TNF-alpha, they are key proteins in inflammatory processes, with the former initiating inflammatory responses through the release of diverse proinflammatory cytokines [6–8] and pyroptosis (a form of programmed lytic cell-death) [9,10] and the latter being a proinflammatory cytokine capable to triggering other inflammation-related proteins (caspase-1 included) [11]. Notice that inflammation has long been accepted as a key component of carcinogenesis because, during inflammation, inflammasomes are potent contributors to cancer progression [12]. In the context of PANC, it has been demonstrated that the inhibition or deletion of components of the NLRP3 inflammasome, such as caspase-1, decreases tumor growth and metastasis in PANC by reprogramming innate and adaptive immunity in the tumor microenvironment [13]. It has also been proven that the inhibition of caspase-1 induced cell death in PANC cells [14]. Additionally, in the case of TNF-alpha, its inhibition led to the diminution of desmoplasia and inflammation to overcome chemoresistance in PANC [15]. In the case of IGF1R, high expression levels of this protein are associated with high tumor grade and poor survival [16], while targeting it inhibited PANC growth and metastasis [17]. In clinical trials, the combination of the well-known anticancer drug gemcitabine with an IGF1R inhibitor (Ganitumab) resulted in a numerical improvement compared to gemcitabine plus placebo [18].

The experimental evidence explained above suggests that the simultaneous inhibition of caspase-1, TNF-alpha, and IGF1R by a multi-target agent could constitute a promising alternative against future PANC treatment. In this sense, finding such multi-target agents can be accelerated utilizing the methodology known as perturbation theory and machine learning (PTML), which allows the integration of different kinds of chemical and biological data [19–23] and has been successfully applied to different drug discovery areas, such as oncology [24–26], neuroscience [27–30], immunology and immunotoxicity [31,32], infectious diseases [33–38], and drug delivery [39]. Considering all the ideas mentioned until now, in this work we establish the theoretical foundations for the rational discovery of multi-target chemicals against caspase-1, TNF-alpha, and IGF1R. Particularly, we report, for the first time, two PTML models based on multi-layer perceptron networks (PTML-MLP) to perform virtual design and prediction of molecules that can simultaneously inhibit not only the aforementioned proteins but also multiple PANC cell lines.

## 2. Materials and Methods

### 2.1. Bioactivity Data and Molecular Descriptors

All the steps necessary for the creation of a PTML-MLP model have been described in detail very recently [37,40]. Therefore, we will focus on the specific aspects of the two PTML-MLP models reported here. Chemical and biological data based on protein inhibition were retrieved from the ChEMBL database [41], while growth inhibition data on PANC cell lines were extracted from the public repository known as Genomics of Drug Sensitivity in Cancer (GDSC) v7.0 [42]. Our dataset contained 3833 different chemicals, each of them experimentally tested by considering at least 1 out of 2 measures of inhibitory activity (*ma*), defined as  $IC_{50}$  (nM)<sub>p</sub> (the concentration required for 50% inhibition of a protein) and  $IC_{50}$  (nM)<sub>c</sub> (the concentration required for a chemical to inhibit cell viability by 50%). Simultaneously, in these experimental assays, each chemical was tested against at least 1 out of 34 biomolecular or cellular targets (*tg*) while involving at least 1 out of 5 types of assay information (*ei*). Notice that each combination of the elements *ma*, *tg*, and *ei*, defined a unique experimental condition, *cj* (denoted as *cj*(*ma*, *tg*, and *ei*)), under which a chemical was assayed. Most of the chemicals were tested in one *cj*. Therefore, after removing entries containing duplicates (only keeping the ones with the lowest inhibition values among the duplicates), as well as those lacking SMILES, activity values, or measurement units, our dataset remained with 9705 statistical cases.

We selected certain cutoff values (Table 1) of inhibitory activity to annotate each case/chemical of our dataset as active ( $IA_i(c_j) = 1$ ) or inactive ( $IA_i(c_j) = -1$ ), with  $IA_i(c_j)$  being a dichotomous variable that indicated the activity of the *i*th case/chemical under a

defined  $c_j$ . These cutoff values were rigorous enough (lower than the 10  $\mu\text{M}$  used in high-throughput screening campaigns) [43] and prevented an unnecessary imbalance between the numbers of cases/chemicals annotated as active and those considered inactive.

For the case of the first PTML-MLP model (*Model 1*), we used the SMILES codes of the 9705 cases/chemicals (already stored in a txt file) and calculated the topological descriptors known as spectral moments of the bond adjacency matrix ( $SM(PP)k$ ) [44–46], atom-based connectivity indices  $X(s)o$  [47,48], atom-based valence connectivity indices  $Xv(s)o$  [48,49], and bond connectivity indices  $e(s)o$  [50]. In this sense, for the topological descriptor  $SM(PP)k$ , the term “ $k$ ” is the  $k$ th power of the bond adjacency matrix and “ $PP$ ” is an atomic physicochemical property, such as hydrophobicity (*Hyd*), polar surface area (*Psa*), molar refractivity (*Mol*), Gasteiger–Marsili charges (*Gas*), and atomic weight (*Atw*). In  $X(s)o$ ,  $Xv(s)o$ , and  $e(s)o$ , the term “ $s$ ” represents the type of subgraph/fragments, such as paths (P), clusters (C), path-clusters (PC), and chains (Ch), while the term “ $o$ ” is the order of the topological descriptor and indicates the number of bonds (without counting bond multiplicity) of each subgraph/fragment. All these topological descriptors were calculated by the software MODESLAB v1.5 [51,52]. Moreover, size-independent topological descriptors (*NTI*) were calculated according to the following formalism:

$$NTI = \frac{TI}{nB} \quad (1)$$

In Equation (1),  $TI$  represents any of the topological descriptors mentioned above and  $nB$  is the numbers of bonds (without counting bond multiplicity) of the molecule.

For the case of the second PTML-MLP model (*Model 2*), the txt file containing the SMILES codes of the 9705 cases/chemicals was manually converted to \*.smi, which was then transformed to \*.sdf (no standardization options were applied) using the software Open Babel v2.4.0 [53]. Then, by using the software QuBiLs-MAS v1.0 (with the \*.sdf file as the input) [54,55], we calculated the descriptors named atom-based local stochastic quadratic indices  $ASqm(x)T$  [54,56], where “ $m$ ” was the  $m$ th power of the atom adjacency matrix and “ $x$ ” was a physicochemical atomic property, such as *Hyd*, electronegativity (*E*), atomic weight (*Aw*), polarizability (*Pol*), *Psa*, and Kupchik’s vertex degree (*Ku*). The term “ $T$ ” referred to the type of atom (aliphatic carbon, aromatic carbon, halogen, carbon in a methyl group, or heteroatoms, such as N, O, S, P, and Se) from which each  $ASqm(x)T$  was calculated.

**Table 1.** Experimental conditions reported in this work.

$ma^a$	Cutoff <sup>b</sup>	$tg^c$	$ei^d$
IC <sub>50</sub> (nM) <sub>p</sub>	≤1100 nM	Caspase-1	B (assay format)
		Caspase-1	B (single protein format)
		Caspase-1	B (cell-based format)
	≤1635 nM	TNF-alpha	B (single protein format)
		TNF-alpha	F (assay format)
		TNF-alpha	B (assay format)
		TNF-alpha	B (cell-based format)
		TNF-alpha	F (cell-based format)
	≤50 nM	IGF1R	B (single protein format)
		IGF1R	B (cell-based format)
		IGF1R	B (assay format)
		IGF1R	F (cell-based format)
		IGF1R	F (assay format)

Table 1. Cont.

<i>ma</i> <sup>a</sup>	Cutoff <sup>b</sup>	<i>tg</i> <sup>c</sup>	<i>ei</i> <sup>d</sup>
		PSN1	F (cell-based format)
		PANC-03-27	F (cell-based format)
		HPAC	F (cell-based format)
		MZ1-PC	F (cell-based format)
		KP-4	F (cell-based format)
		KP-2	F (cell-based format)
		PA-TU-8988T	F (cell-based format)
		Capan-2	F (cell-based format)
		MIA-PaCa-2	F (cell-based format)
		CFPAC-1	F (cell-based format)
		PANC-10-05	F (cell-based format)
		BxPC-3	F (cell-based format)
IC <sub>50</sub> (nM) <sup>c</sup>	≤6449.735 nM	SUIT-2	F (cell-based format)
		KP-1N	F (cell-based format)
		HuP-T4	F (cell-based format)
		SW1990	F (cell-based format)
		PL18	F (cell-based format)
		QGP-1	F (cell-based format)
		HuP-T3	F (cell-based format)
		SU8686	F (cell-based format)
		PL4	F (cell-based format)
		PA-TU-8902	F (cell-based format)
		PANC-02-03	F (cell-based format)
		DAN-G	F (cell-based format)
		CAPAN-1	F (cell-based format)
		PANC-08-13	F (cell-based format)
		HPAF-II	F (cell-based format)
		KP-3	F (cell-based format)
		YAPC	F (cell-based format)
		AsPC-1	F (cell-based format)
		PANC-04-03	F (cell-based format)

<sup>a</sup> Measure of biological activity; IC<sub>50</sub> (nM)<sub>p</sub> is the concentration required for 50% inhibition of a protein, while IC<sub>50</sub> (nM)<sub>c</sub> is the concentration required for a chemical to inhibit cell viability by 50%. <sup>b</sup> Value of activity from which a molecule was labeled and considered as active ( $I_{Ai}(cj) = 1$ ). <sup>c</sup> Refers to the targets (either a protein or a PANC cell line). <sup>d</sup> Information related to the diverse experimental assays. Here, each annotation is a combination of the columns “assay type” (first letter) and “BioAssay Ontology” (phrase between parentheses), which are reported in any ChEMBL file containing bioactivity data. Each assay involving a PANC cell line was annotated as “F (cell-based format)”.

To consider both the structure of any case/chemical and the experimental condition, *cj*, under which that case/chemical was tested, we applied a two-step approach, known as Box–Jenkins, which is the key aspect accounting for the great success of the PTML models [24–37,57–62]:

$$avg[GTI]_{cj} = \frac{1}{n(cj)} \times \sum_{i=1}^{n(cj)} GTI_i \quad (2)$$

In Equation (2),  $GTI$  refers to any of the molecular descriptors discussed above, i.e.,  $SM(PP)k$ ,  $X(s)o$ ,  $Xv(s)o$ ,  $e(s)o$ ,  $NTI$ , and  $ASqm(x)T$ . The meanings of the terms  $avg(GTI)_{cj}$  and  $n(cj)$  have already been explained in detail in a recent work [63]. This means that Equation (2) was applied to each element of the experimental conditions  $cj$  (that is,  $ma$ ,  $tg$ , and  $ei$ ) separately. Then, the second step of the Box–Jenkins approach was applied:

$$D[GTI]_{cj} = \left( \frac{GTI - avg[GTI]_{cj}}{std(GTI)} \right) \cdot \sqrt{ps(cj)} \quad (3)$$

In Equation (3),  $D(GTI)_{cj}$  is a descriptor that measures how much a chemical structurally and physicochemically deviates from a group of chemicals assayed by considering the same element of the experimental condition  $cj$ . On the other hand,  $std(GTI)$  is the standard deviation calculated from the  $GTI$  values; only chemicals in the training set were considered for the calculation of  $std(GTI)$ . Last,  $ps(cj)$  represents the *a priori* probability of finding a case/chemical annotated as active by considering a defined element of  $cj$ . Thus,  $ps(cj)$  was calculated as follows:

$$ps(cj) = \frac{n(cj)}{N_T(cj)} \quad (4)$$

In Equation (4),  $n(cj)$  has been defined in Equation (2) and represents the number of cases/chemicals assayed by considering the same element of the experimental condition  $cj$  [63], which were annotated as active (in the training set). Similarly,  $N_T(cj)$ , considering the same condition  $cj$ , represents the total number of cases/chemicals in the training set. We would like to emphasize that, as in the case of Equation (2), Equations (3) and (4) were applied to each element of the experimental condition  $cj$  ( $ma$ ,  $tg$ , and  $ei$ ) separately.

## 2.2. PTML Modeling, Applicability Domain, Descriptor Interpretation, Fragments, and Virtual Design

The creation and application of the two PTML-MLP models developed in this work involved steps such as splitting the dataset in the training and test series, selecting the most suitable  $D(GTI)_{cj}$  descriptors using the software IMMAN v1.0 [64], analysis of the correlations among the  $D(GTI)_{cj}$  descriptors via the Pearson correlation coefficient ( $PCC$ ) [65], generation of the models using the program STATISTICA v13.5.0.17 [66], analysis of the applicability domain of each PTML-MLP model, interpretation of the  $D(GTI)_{cj}$  descriptors, selection of suitable molecular fragments, and virtual design. All these steps have been described comprehensively in seminal works [37,63,67–69]. In any case, when selecting the best  $D(GTI)_{cj}$  descriptors to subsequently build the PTML-MLP models, the mutual information differential Shannon's entropy (MI-DSE) [70] and the Jeffreys information [71,72] were applied as criteria for descriptor selection; such criteria permitted the selection of at least one  $D(GTI)_{cj}$  descriptor per each element of the experimental condition  $cj$ , which was a mandatory condition to develop the PTML-MLP models. When estimating the correlations via  $PCC$ , the interval  $-0.7 < PCC < 0.7$  was used as a criterion of a lack of redundancy among the  $D(GTI)_{cj}$  descriptors. We analyzed the global performance of the PTML-MLP models by relying on measures such as sensitivity ( $Sn(\%)$ ), specificity ( $Sp(\%)$ ), accuracy ( $Ac(\%)$ ), and the Matthews' correlation coefficient ( $MCC$ ) [73]. However, when choosing the most appropriate measure we determined the values of the local counterparts of ( $Sn(\%)$ ) and ( $Sp(\%)$ ), i.e., the local sensitivities ( $Sn(\%)_{ma}$ , ( $Sn(\%)_{tg}$ , and ( $Sn(\%)_{ei}$ ), as well as the local specificities ( $Sp(\%)_{ma}$ , ( $Sp(\%)_{tg}$ , and ( $Sp(\%)_{ei}$ ). Notice that these six local statistical indices depended on specific elements of the experimental condition  $cj$  ( $ma$ ,  $tg$ , and  $ei$ ) and we chose the PTML-MLP models displaying the highest values of the aforementioned local metrics.

### 3. Results and Discussion

#### 3.1. PTML-MLP Models

We found that *Model 1* had the notation MLP 15-45-2, which means that this model was based on a multilayer perceptron network with 15 nodes in the input layer (number of  $D(GTI)_{cj}$  descriptors present in the model), 45 neurons in the hidden layer, and two values of the categorical variable of inhibitory activity  $I_{Ai}(cj)$  were predicted in the output layer, i.e., active ( $I_{Ai}(cj) = 1$ ) and inactive ( $I_{Ai}(cj) = -1$ ). The same deduction can be made for *Model 2* whose notation was MLP 14-45-2. A summary of the  $D(GTI)_{cj}$  descriptors present in each PTML-MLP model appears in Table 2, while information regarding the chemical and biological data used to build such models appears in Supplementary Materials S1 and S2, respectively.

**Table 2.** Molecular descriptors of the type  $D(GTI)_{cj}$  present in the PTML-MLP models.

Model <sup>a</sup>	Symbology <sup>b</sup>	Code <sup>c</sup>	Concept
	$D(NSM(Hyd)3)_{ma}$	DT01	Deviation of the normalized spectral moment of order 3 based on hydrophobicity-weighted bonds.
	$D(NXv(P)4)_{ma}$	DT02	Deviation of the normalized Kier–Hall (valence) connectivity index involving only path-based subgraphs of order 4.
	$D(Ne(P)1)_{ma}$	DT03	Deviation of the normalized edge (bond) connectivity index involving only path-based subgraphs of order 1.
	$D(Ne(P)2)_{ma}$	DT04	Deviation of the normalized edge (bond) connectivity index involving only path-based subgraphs of order 2.
	$D(Ne(Ch)6)_{ma}$	DT05	Deviation of the normalized edge (bond) connectivity index involving only chain-based subgraphs of order 6.
	$D(SM(Hyd)7)_{tg}$	DT06	Deviation of the spectral moment of order 7 based on hydrophobicity-weighted bonds.
	$D(e(Ch)5)_{tg}$	DT07	Deviation of the edge (bond) connectivity index involving only chain-based subgraphs of order 5.
<i>Model 1</i>	$D(NSM(Psa)1)_{tg}$	DT08	Deviation of the normalized spectral moment of order 1 based on bonds weighted by the polar surface area.
	$D(NSM(Gas)3)_{tg}$	DT09	Deviation of the normalized spectral moment of order 3 based on bonds weighted by the Gasteiger–Marsili charges.
	$D(NXv(P)1)_{tg}$	DT10	Deviation of the normalized Kier–Hall (valence) connectivity index involving only path-based subgraphs of order 1.
	$D(Xv(Ch)6)_{ei}$	DT11	Deviation of the Kier–Hall (valence) connectivity index involving only chain-based subgraphs of order 6.
	$D(NSM(Hyd)1)_{ei}$	DT12	Deviation of the normalized spectral moment of order 1 based on hydrophobicity-weighted bonds.
	$D(NSM(Mol)1)_{ei}$	DT13	Deviation of the normalized spectral moment of order 1 based on bonds weighted by the molar refractivity.
	$D(Ne(P)5)_{ei}$	DT14	Deviation of the normalized edge (bond) connectivity index involving only path-based subgraphs of order 5.
	$D(Ne(PC)6)_{ei}$	DT15	Deviation of the normalized edge (bond) connectivity index involving only path-cluster subgraphs of order 6.

Table 2. Cont.

Model <sup>a</sup>	Symbology <sup>b</sup>	Code <sup>c</sup>	Concept
Model 2	$D(ASq4(Hyd)G)ma$	DQ01	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the halogens and their neighbor atoms located at the topological distance of 4.
	$D(ASq3(Hyd)Y)ma$	DQ02	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the heteroatoms (N, O, S, P, and Se) and their neighbor atoms located at the topological distance of 3.
	$D(ASq4(Hyd)Y)ma$	DQ03	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the heteroatoms (N, O, S, P, and Se) and their neighbor atoms located at the topological distance of 4.
	$D(ASq2(E)Y)ma$	DQ04	Deviation of the stochastic atom-based local quadratic index weighted by the electronegativity of the heteroatoms (N, O, S, P, and Se) and their neighbor atoms located at the topological distance of 2.
	$D(ASq1(Psa)Y)ma$	DQ05	Deviation of the stochastic atom-based local quadratic index weighted by the polar surface area of the heteroatoms (N, O, S, P, and Se) and their neighbor atoms located at the topological distance of 1.
	$D(ASq1(Aw)C)tg$	DQ06	Deviation of the stochastic atom-based local quadratic index weighted by the atomic weight of the aliphatic carbons and their neighbor atoms located at the topological distance of 1.
	$D(ASq0(Ku)G)tg$	DQ07	Deviation of the stochastic atom-based local quadratic index (order 0) weighted by the Kupchik's vertex degree of the halogens in a molecule.
	$D(ASq4(Psa)Y)tg$	DQ08	Deviation of the stochastic atom-based local quadratic index weighted by the polar surface area of the heteroatoms (N, O, S, P, and Se) and their neighbor atoms located at the topological distance of 4.
	$D(ASq1(Hyd)G)ei$	DQ09	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the halogens and their neighbor atoms located at the topological distance of 1.
	$D(ASq2(Hyd)G)ei$	DQ10	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the halogens and their neighbor atoms located at the topological distance of 2.
	$D(ASq2(Aw)G)ei$	DQ11	Deviation of the stochastic atom-based local quadratic index weighted by the atomic weight of the halogens and their neighbor atoms located at the topological distance of 2.
	$D(ASq1(Hyd)M)ei$	DQ12	Deviation of the stochastic atom-based local quadratic index weighted by the hydrophobicity of the aliphatic carbons (only methyl groups) and their neighbor atoms located at the topological distance of 1.
	$D(ASq1(Ku)M)ei$	DQ13	Deviation of the stochastic atom-based local quadratic index weighted by the Kupchik's vertex degree of the aliphatic carbons (only methyl groups) and their neighbor atoms located at the topological distance of 1.
	$D(ASq0(Hyd)Y)ei$	DQ14	Deviation of the stochastic atom-based local quadratic index (order 0) weighted by the hydrophobicity heteroatoms (N, O, S, P, and Se) in a molecule.

<sup>a</sup> Model 1, first PTML-MLP model, which contains the first 15  $D(GTI)cj$  descriptors shown in this table; Model 2, second PTML-MLP model, which contains the remaining 14  $D(GTI)cj$  descriptors shown in this table. <sup>b</sup> Molecular descriptors of the type  $D(GTI)cj$  with endings on "ma" consider both the molecular structure and the measure of inhibitory activity. Those with the ending "tg" depend on the molecular structure and the biological target (either a protein or a PANC cell line). Finally,  $D(GTI)cj$  descriptors with the ending "ei" characterize the molecular structure and information on the diverse experimental assays. <sup>c</sup> Codes were used to abbreviate the representation of the  $D(GTI)cj$  descriptors.

Here, Model 1 correctly classified 6312 out of the 7283 cases/chemicals in the training set (internal quality), which means  $Ac(\%) = 86.67\%$ . In the test set (predictive power), Model 1 satisfactorily classified 2006 out of 2422 cases/chemicals, with  $Ac(\%) = 82.82\%$ . In the case of Model 2, similar results were achieved; this PTML-MLP model had  $Ac(\%)$

values of 86.49% (6299 out of the 7283 cases/chemicals were correctly classified) and 81.75% (1980 out of 2422 cases/chemicals were correctly classified) for the training and test sets, respectively. Moreover, Table 3 shows that the two PTML-MLP models have high  $Sp(\%)$  values, surpassing 80% in both the training and test set. In addition, for comparison, in Table 3 we have reported the values of the different statistical indices for the linear counterparts of our two PTML-MLP models. Such linear models were based on the technique known as linear discriminant analysis (LDA), and they respectively used the same variables (molecular descriptors of the type  $D(GTI)cj$ ) from which our two PTML-MLP models were constructed.

**Table 3.** Statistical indices demonstrating the performances of the two PTML-MLP models.

SYMBOLS <sup>a,b</sup>	Model 1		Model 2	
	Training Set	Test Set	Training Set	Test Set
$N_{Active}$	3010	1001	3010	1001
$CCC_{Active}$	2495 (1293)	799 (447)	2486 (1084)	785 (352)
$Sn(\%)$	82.89% (42.96%)	79.82% (44.66%)	82.59% (36.01%)	78.42% (35.16%)
$N_{Inactive}$	4273	1421	4273	1421
$CCC_{Inactive}$	3817 (3693)	1207 (1219)	3813 (3625)	1195 (1194)
$Sp(\%)$	89.33% (86.43%)	84.94% (85.78%)	89.23% (84.84%)	84.10% (84.03%)
MCC	0.724 (0.331)	0.646 (0.338)	0.721 (0.241)	0.624 (0.222)

<sup>a</sup>  $N_{Active}$ , Number of chemicals/cases annotated as active;  $N_{Inactive}$ , Number of chemicals/cases designated as inactive;  $CCC_{Active}$ , Number of chemicals/cases correctly classified/predicted as active;  $CCC_{Inactive}$ , Number of chemicals/cases correctly classified/predicted as inactive;  $Sn(\%)$ , Sensitivity (percentage of chemicals/cases correctly classified as active);  $Sp(\%)$ , Specificity (percentage of chemicals/cases properly classified as inactive); MCC, Refers to the Matthews' correlation coefficient. <sup>b</sup> Values between parentheses correspond to models derived from the technique known as linear discriminant analysis (LDA).

From the analysis of the classification results in Table 3, it can be seen that each of our two PTML-MLP models outperforms its corresponding LDA counterpart. This suggests that the relationship between the measures of anti-PANC activity and the chemical structure of the molecules in the present dataset is modeled better with the use of non-linear machine learning algorithms, as in the case of our two PTML-MLP models.

Continuing with the results depicted in Table 3 for our PTML-MLP models, we can observe that for the case statistical index  $Sn(\%)$ , the values were higher than 80% in the training set while remaining above 75% in the test set. Moreover, in both PTML-MLP models, there is a strong convergence between the observed and the predicted values of the categorical variable of inhibitory activity  $IAi(cj)$  since the statistical metric MCC is closer to 1 than to 0 (random classifier) or  $-1$  (completely erroneous prediction). Altogether, the results from Table 3 indicate that *Model 1* classified/predicted the current data slightly better than *Model 2*; however, *Model 2*, using one less  $D(GTI)cj$  descriptor than *Model 1* achieved similar performance.

We went deeper and, thus, analyzed the local metrics derived from  $Sn(\%)$  and  $Sp(\%)$ . In this sense, *Model 1* exhibited values in the interval 69.64–99.21% for  $(Sn(\%))ma$ ,  $(Sn(\%))tg$ ,  $(Sn(\%))ei$ ,  $(Sp(\%))ma$ ,  $(Sp(\%))tg$ , and  $(Sp(\%))ei$  in the training set. The only exception was  $(Sn(\%))ei = 33.33\%$  for the assay information labeled as “F (assay format)”. In the test set, the behavior of *Model 1* was very similar to the training set, as the aforementioned local metrics were in the range of 64.29–100%. In this test set, *Model 1* maintained the same exception as in the training set and added the PANC cell line MIA-PaCa-2 with  $(Sn(\%))tg = 55.56\%$  and the assay information labeled as “B (assay format)” with  $(Sp(\%))ei = 57.14\%$ . In the case of *Model 2*, values were between 65.22% and 97.14% in the training set; a relatively low performance was associated with the PANC cell line PSN1, exhibiting  $(Sn(\%))tg = 57.29\%$  and the assay information annotated as “F (assay format)” displaying  $(Sn(\%))ei = 44.44\%$ . In the test set, *Model 2* achieved the interval 60.87–100%

for the 6 local metrics mentioned above, with the exceptions being the labels of assay information “B (assay format)” with  $(Sp(\%))_{ei} = 53.97\%$  and assay information labeled as “F (assay format)” with  $(Sn(\%))_{ei} = 38.46\%$ . We would like to emphasize that most of the values reported in this work for  $(Sn(\%))_{ma}$ ,  $(Sn(\%))_{tg}$ ,  $(Sn(\%))_{ei}$ ,  $(Sp(\%))_{ma}$ ,  $(Sp(\%))_{tg}$ , and  $(Sp(\%))_{ei}$  were above 70%, which confirms the great capabilities of both PTML-MLP models to classify/predict complex biological data focused on anti-PANC activity by considering the 44 different experimental conditions,  $c_j$  (as depicted in Table 1), employed in this study. All the details regarding the classification results of *Model 1* and *Model 2* can be found in Supplementary Materials S3 and S4, respectively.

Regarding the reliability of the predictions, we determined the AD of both PTML-MLP models according to the descriptor space approach (Supplementary Material S5), and, therefore, we calculated the total scores of the applicability domain (TSAD). For *Model 1*, only the cases/chemicals with TSAD = 15 were considered to be within the AD, with the number fifteen being equivalent to the number of  $D(GTI)c_j$  descriptors present in *Model 1*. Only 2 of the 9705 cases/chemicals were found to be outside the AD of *Model 1* (TSAD < 15) and they both belonged to the test set. For *Model 2*, the ideal value TSAD = 14 indicated the ability of case/chemical to fall within the AD. In this model, only 8 of the 9705 cases/chemicals had TSAD < 14, thus remaining outside of the AD of *Model 2*.

Last, we would like to highlight the limitations and strengths of our two PTML-MLP models. A key limitation of our PTML-MLP models is their inability to correctly predict the entire dataset employed in this work. This means that the  $D(GTI)c_j$  descriptors used to build the PTML-MLP models cannot characterize the whole complexity and molecular diversity in the present dataset. Another limitation is the one associated with the machine learning algorithm since none of the current algorithms are capable of encoding enough accurate chemical information. At least one of these two aspects is generalized to all computational models reported in the scientific literature. Consequently, in a virtual screening scenario, our PTML-MLP models will perform accurate predictions to some extent.

In any case, the strengths of our PTML-MLP models outweigh the aforementioned limitation. First, our PTML-MLP models are the first two models reported to date that can predict anti-PANC activity by simultaneously considering different mechanisms of action and multiple experimental conditions (involving dissimilar assay protocols) on the inhibition of both PANC-related proteins and PANC cell lines. This highlights the potentialities and wide applications of the PTML methodology in integrating chemical and biological data in the context of oncology research [24–26,74–79]. Second, our two PTML-MLP models are highly interpretable, physicochemically and structurally describing the data in terms of suitable molecular fragments (see Section 3.2), which can be useful to both computational and experimental chemists when selecting 2D pharmacophores. Third, by interpreting our PTML-MLP models, it is possible to design new anti-PANC molecules (see Section 3.3). Last, in contrast to most computational models, whose complex algorithms involve remarkable consumption of time and financial resources, our PTML-MLP models are cost-efficient; it only took two hours to create each of them when working on a portable computer (12 GB RAM).

### 3.2. Physicochemical and Structural Meanings of the Molecular Descriptors

Here, we are providing the physicochemical and structural interpretations of the  $D(GTI)c_j$  descriptors in the PTML-MLP models. To support such interpretations, we employed certain graphics that illustrated the significance of each  $D(GTI)c_j$  descriptor in the form of sensitivity values (SV) while extracting information on the propensities (increment or diminution) in the value of each  $D(GTI)c_j$  descriptor [37,63,80]. Such propensities indicate how to vary the values of the  $D(GTI)c_j$  descriptors to enhance the desired activity, which, in our case, is the improvement of the inhibitory activity against both the PANC-related proteins and the PANC cell lines.

## 3.2.1. First PTML-MLP Model (Model 1)

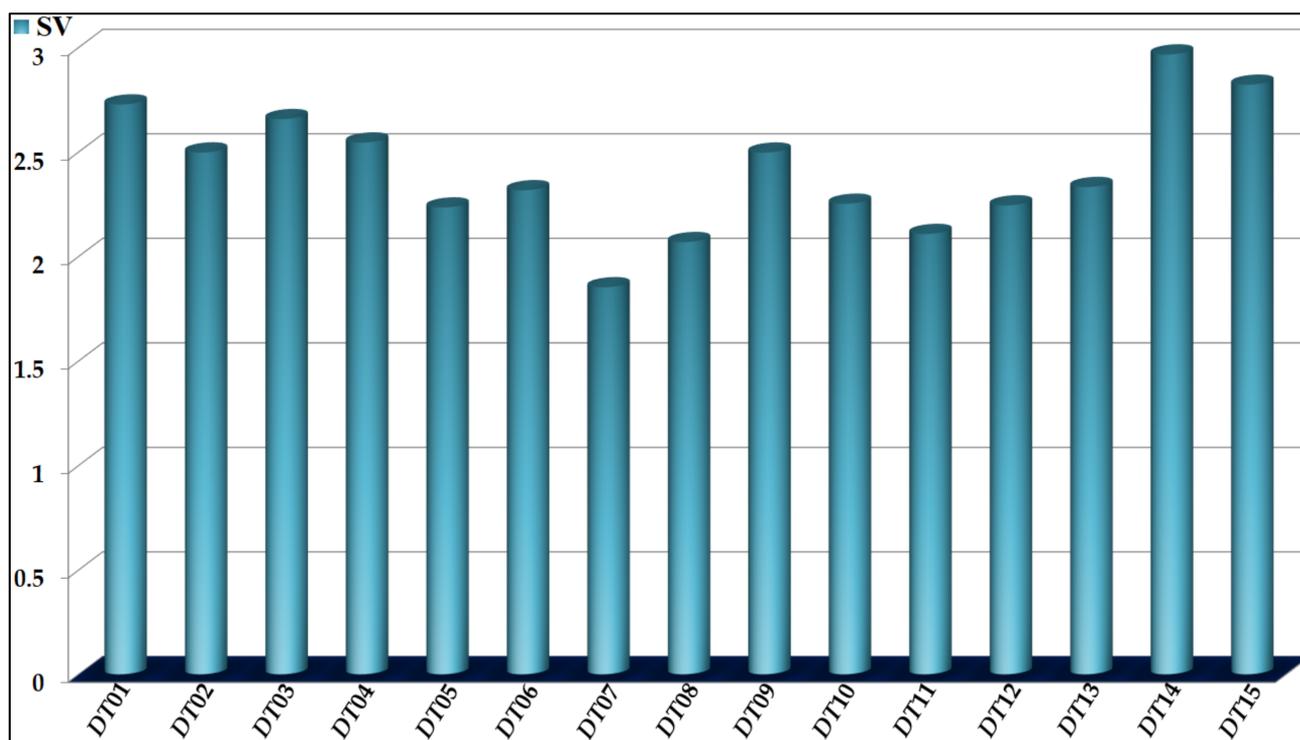
The propensities of the  $D(GTI)_{cj}$  descriptors in *Model 1* are reported in Table 4, while the relative importance of each of them (measured by  $SV$ ) is depicted in Figure 1.

In *Model 1*, we have six  $D(GTI)_{cj}$  descriptors derived from the so-called spectral moments of the bond adjacency matrix, which denote how much the physicochemical properties are distributed/concentrated throughout the chemical structure of a molecule [44–46,81–83]. These  $D(GTI)_{cj}$  descriptors are  $DTI01$ ,  $DTI06$ ,  $DTI08$ ,  $DTI09$ ,  $DTI12$ , and  $DTI13$  and among the most significant descriptors in *Model 1*, they rank third, ninth, fourteenth, seventh, eleventh, and eighth, respectively.

**Table 4.** Molecular descriptors of the type  $D(GTI)_{cj}$  present in the first PTML-MLP model (*Model 1*) and their relative propensities.

Codes <sup>a</sup>	Descriptors	CLASS-BASED MEANS <sup>b</sup>		Propensity <sup>c</sup>
		Active	Inactive	
$DTI01$	$D(NSM(Hyd)3)_{ma}$	$-2.3485 \times 10^{-2}$	$1.0631 \times 10^{-1}$	Decrease
$DTI02$	$D(NXv(P)4)_{ma}$	$6.6912 \times 10^{-3}$	$5.5922 \times 10^{-3}$	Increase
$DTI03$	$D(Ne(P)1)_{ma}$	$-3.2309 \times 10^{-4}$	$8.6125 \times 10^{-2}$	Decrease
$DTI04$	$D(Ne(P)2)_{ma}$	$-4.7514 \times 10^{-2}$	$1.4992 \times 10^{-1}$	Decrease
$DTI05$	$D(Ne(Ch)6)_{ma}$	$-4.1335 \times 10^{-2}$	$1.7654 \times 10^{-1}$	Decrease
$DTI06$	$D(SM(Hyd)7)_{tg}$	$3.1706 \times 10^{-2}$	$-2.5668 \times 10^{-1}$	Increase
$DTI07$	$D(e(Ch)5)_{tg}$	$5.2858 \times 10^{-3}$	$-4.9427 \times 10^{-2}$	Increase
$DTI08$	$D(NSM(Psa)1)_{tg}$	$2.2379 \times 10^{-2}$	$-1.9481 \times 10^{-2}$	Increase
$DTI09$	$D(NSM(Gas)3)_{tg}$	$2.8122 \times 10^{-2}$	$-1.3897 \times 10^{-1}$	Increase
$DTI10$	$D(NXv(P)1)_{tg}$	$-1.2919 \times 10^{-2}$	$1.1367 \times 10^{-1}$	Decrease
$DTI11$	$D(Xv(Ch)6)_{ei}$	$-4.9125 \times 10^{-3}$	$-7.1198 \times 10^{-2}$	Increase
$DTI12$	$D(NSM(Hyd)1)_{ei}$	$-5.5366 \times 10^{-2}$	$2.3959 \times 10^{-1}$	Decrease
$DTI13$	$D(NSM(Mol)1)_{ei}$	$-4.2905 \times 10^{-2}$	$2.2042 \times 10^{-1}$	Decrease
$DTI14$	$D(Ne(P)5)_{ei}$	$-1.2401 \times 10^{-2}$	$-5.6792 \times 10^{-2}$	Increase
$DTI15$	$D(Ne(PC)6)_{ei}$	$3.4465 \times 10^{-2}$	$-2.3273 \times 10^{-1}$	Increase

<sup>a</sup> Symbols of the different molecular descriptors of the type  $D(GTI)_{cj}$  in *Model 1* as represented in Table 2. <sup>b</sup> Average values of each  $D(GTI)_{cj}$  descriptor by considering the active and inactive categories. <sup>c</sup> Relative tendency of a molecular descriptor to vary (increase or decrease) its value, resulting in a simultaneous enhancement of the inhibitory activity against PANC-related proteins (caspase-1, TNF-alpha, and IGF1R) and the PANC cell lines.



**Figure 1.** The different  $D(GTI)c_j$  descriptors and their relative significances in *Model 1*.

Three of these  $D(GTI)c_j$  descriptors are derived from the spectral moments and are focused on hydrophobicity, according to the Ghose–Crippen approach [84]. Thus,  $DTI01$  characterizes the decrease of the hydrophobicity in regions containing three-membered rings and fragments where one atom is attached to three other non-hydrogen atoms. Therefore, the presence of urea, carbamate, and carboxamide groups, as well as portions with secondary alcohols and amines, methylene moieties, and aziridine and oxirane rings favor a decrease in the value of  $DTI01$ . The same fragments mentioned in the case of  $DTI01$ , will also desirably decrease the value of  $DTI12$ ; however, the latter of these  $D(GTI)c_j$  descriptors will impact a molecule more globally. Another hydrophobicity-based descriptor is  $DTI06$ , which involves the increment of the hydrophobicity in regions in fragments containing seven bonds or less. The presence of three-membered rings and groups/moieties where an atom is bonded to four other non-hydrogen atoms (e.g., quaternary carbons, sulfonamides, and phosphorus-based functional groups, such as phosphine oxides, phosphonates, and phosphates) will increase the value of  $DTI06$ . At the same time,  $DTI08$  describes the increase of the polar surface area, which can be achieved by increasing the number the functional groups (mainly based on nitrogen and oxygen atoms) capable of forming hydrogen bonds. On the other hand,  $DTI09$  implies the augmentation of electronegative atoms in three-membered rings and groups where one atom is attached to two other non-hydrogen atoms. The functional groups described by  $DTI01$  also favor  $DTI09$ , but the latter characterizes the structure in terms of the distribution of partial charges, also including the increase of halogens (particularly fluor). The last  $D(GTI)c_j$  descriptor derived from the spectral moments is  $DTI13$ , which gives information regarding the diminution of the global polarizability in the molecules, thus favoring the presence of nitrogen, oxygen, and fluorine atoms, while diminishing the presence of aromatic rings or alkene moieties, as well as the functional groups attached to them.

In *Model 1*, we also have  $D(GTI)c_j$  descriptors derived from atom-based connectivity indices, and, therefore, these  $D(GTI)c_j$  descriptors measure the contribution of different fragments to the molecular accessibility [85,86], i.e., the ability of certain regions of a given molecule to interact with the surrounding chemical environment (e.g., with molecules such as water, amino acids in the cavity of a protein, and different components of the

membrane PANC cell line). The  $D(GTI)_{cj}$  descriptors are  $DTI02$ ,  $DTI10$ , and  $DTI11$ , and, among the most influential, they ranked sixth, tenth, and thirteenth, respectively. Thus, in *Model 1*,  $DTI02$  describes the increment of the molecular accessibility, mainly in linear (non-ramified) fragments containing at least one atom different from carbon, hydrogen, and boron. In contrast with  $DTI02$ ,  $DTI10$  measures the decrease of the average accessibility area of a molecule, which can be achieved by limiting the number of sulfur and halogen atoms (except fluor). In the case of  $DTI11$ , this descriptor indicates that augmenting the accessibility (and, therefore, the number) of six-membered rings will play a favorable role in enhancing the inhibitory activity against both the PANC-related proteins and the PANC cell lines.

Last, we have  $D(GTI)_{cj}$  descriptors which encompass information content similar to that of the edge (bond) connectivity indices and, therefore, behave as measures of contribution to the molecular/molar volume [50,87,88]. Such  $D(GTI)_{cj}$  descriptors under analysis are  $DTI03$ ,  $DTI04$ ,  $DTI05$ ,  $DTI07$ ,  $DTI14$ , and  $DTI15$ , and they have been considered as the fourth, fifth, twelfth, fifteenth, first, and second most significant descriptors in *Model 1*, respectively. Thus,  $DTI03$  is a direct measure of the global volume of a molecule, expressing the decrease of this physicochemical property. This has important implications because diminishing the molecular volume means that a molecule will be able to better fit in the binding pocket of the PANC-related proteins while also allowing the molecule to be small enough to permeate the cellular membranes [89]. Decreasing the number of atoms in a molecule will favorably decrease the value of  $DTI03$ . A similar effect is expected to take place by diminishing the value of  $DTI04$ . The only difference is that  $DTI04$  focuses on fragments formed by two bonds (without counting bond multiplicity). In the case of  $DTI05$ , this expresses the decrease of the volume in regions containing six-membered rings. We should highlight that this descriptor constrains  $DTI11$ , which was explained before. In the end, at least two six-membered rings can be part of the structure of a molecule, but these rings should have at least two substitutions while the presence of nitrogen atoms in these rings is a favorable factor. The descriptor  $DTI07$  characterizes the increase of the molecular volume due to the presence of five-membered rings. The two most important  $D(GTI)_{cj}$  descriptors ( $DTI14$  and  $DTI15$ ) indicate the increase of the average volume by decreasing the number of atoms in a molecule while increasing the number of regions without branching in that molecule. Particularly,  $DTI14$  measures this effect in five-bond linear fragments while  $DTI15$  focuses on six-bond fragments where the presence of a six-membered ring is also important.

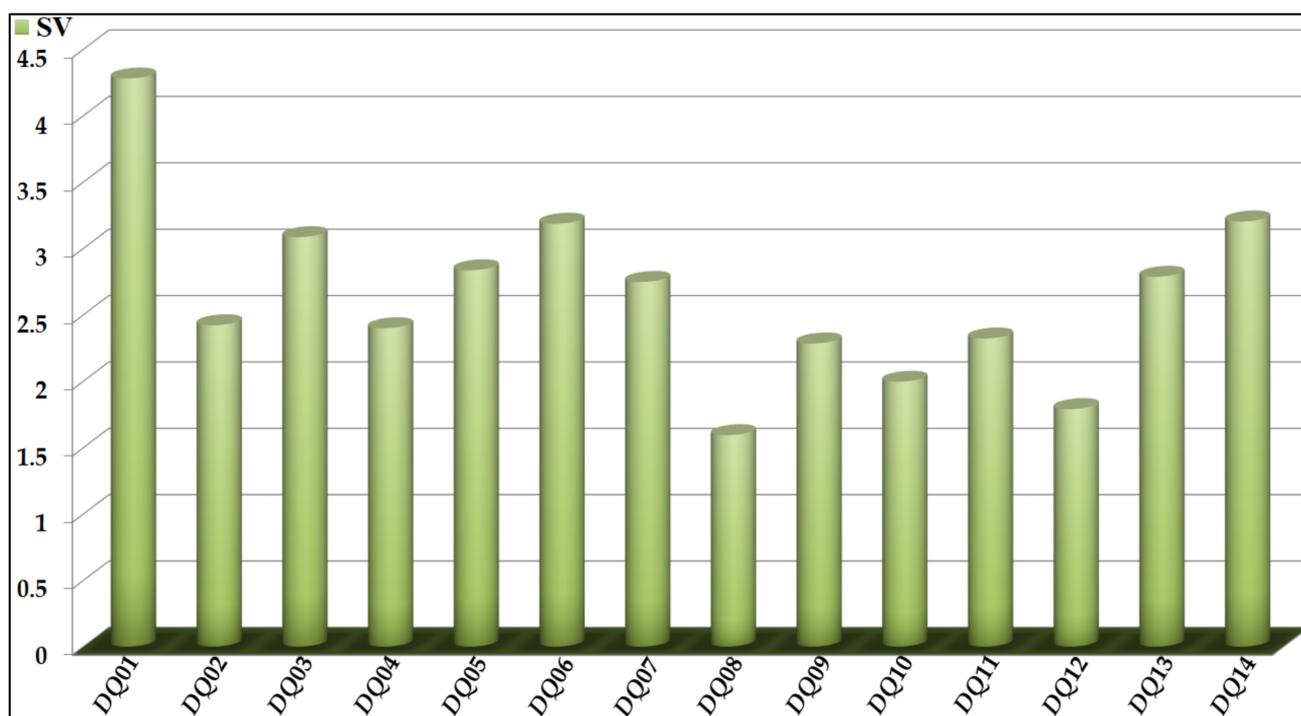
### 3.2.2. Second PTML-MLP Model (Model 2)

As in the case of *Model 1*, data regarding each  $D(GTI)_{cj}$  descriptor propensity and significance in *Model 2* can be found in Table 5 and Figure 2, respectively. The  $D(GTI)_{cj}$  descriptors used in *Model 2* were derived from the atom-based local stochastic quadratic indices [54,56,90], which means that they characterize the distribution of different physicochemical properties of diverse fragments containing specific types of atoms [37,90]. In *Model 2*, seven of the fourteen  $D(GTI)_{cj}$  descriptors are based on hydrophobicity, which indicates the paramount importance of this physicochemical property. These  $D(GTI)_{cj}$  descriptors are  $DQI01$ ,  $DQI02$ ,  $DQI03$ ,  $DQI09$ ,  $DQI10$ ,  $DQI12$ , and  $DQI14$  and we would like to highlight that, these  $D(GTI)_{cj}$  descriptors consider the multiplication of the hydrophobic contribution of a specific type of atom and the hydrophobic contribution of other non-hydrogen atoms situated at the given topological distance (number of bonds between any two atoms without counting bond multiplicity). Hence, instead of referring to an atomic hydrophobicity, we will use the term “joint hydrophobic contribution”.

**Table 5.** Class-based means and relative propensities of the  $D(GTI)_{cj}$  descriptors present in the second PTML-MLP model (*Model 2*).

Codes <sup>a</sup>	Descriptors	CLASS-BASED MEANS <sup>b</sup>		Propensity <sup>c</sup>
		Active	Inactive	
DQI01	$D(ASq4(Hyd)G)ma$	$-5.4750 \times 10^{-3}$	$-4.6447 \times 10^{-3}$	Decrease
DQI02	$D(ASq3(Hyd)Y)ma$	$2.6348 \times 10^{-2}$	$-1.8596 \times 10^{-1}$	Increase
DQI03	$D(ASq4(Hyd)Y)ma$	$3.5673 \times 10^{-2}$	$-1.3724 \times 10^{-1}$	Increase
DQI04	$D(ASq2(E)Y)ma$	$5.0259 \times 10^{-2}$	$-2.2849 \times 10^{-1}$	Increase
DQI05	$D(ASq1(Psa)Y)ma$	$3.8557 \times 10^{-3}$	$6.3789 \times 10^{-2}$	Decrease
DQI06	$D(ASq1(Aw)C)tg$	$5.9744 \times 10^{-2}$	$-3.4756 \times 10^{-1}$	Increase
DQI07	$D(ASq0(Ku)G)tg$	$3.9557 \times 10^{-3}$	$1.1249 \times 10^{-1}$	Decrease
DQI08	$D(ASq4(Psa)Y)tg$	$3.1247 \times 10^{-2}$	$-1.2646 \times 10^{-1}$	Increase
DQI09	$D(ASq1(Hyd)G)ei$	$-1.1160 \times 10^{-2}$	$4.0920 \times 10^{-2}$	Decrease
DQI10	$D(ASq2(Hyd)G)ei$	$5.8725 \times 10^{-4}$	$3.3683 \times 10^{-2}$	Decrease
DQI11	$D(ASq2(Aw)G)ei$	$9.2750 \times 10^{-3}$	$1.9038 \times 10^{-2}$	Decrease
DQI12	$D(ASq1(Hyd)M)ei$	$2.6208 \times 10^{-2}$	$-1.7829 \times 10^{-1}$	Increase
DQI13	$D(ASq1(Ku)M)ei$	$2.8400 \times 10^{-2}$	$-3.0151 \times 10^{-1}$	Increase
DQI14	$D(ASq0(Hyd)Y)ei$	$1.3478 \times 10^{-2}$	$-2.1995 \times 10^{-2}$	Increase

<sup>a</sup> Symbols of the different  $D(GTI)_{cj}$  descriptors in the second PTML-MLP model according to Figure 2. <sup>b</sup> Average values of each  $D(GTI)_{cj}$  descriptor by considering the active and inactive categories. <sup>c</sup> Relative tendency of each  $D(GTI)_{cj}$  descriptor to vary (increase or decrease) its value, resulting in a simultaneous enhancement of the inhibitory activity against PANC-related proteins (caspase-1, TNF-alpha, and IGF1R) and the PANC cell lines.

**Figure 2.** Relative importance of the different  $D(GTI)_{cj}$  descriptors in *Model 2*.

In terms of significance in *Model 2*, the  $D(GTI)_{cj}$  descriptors DQI01, DQI02, DQI03, DQI09, DQI10, DQI12, and DQI14 rank first, eighth, fourth, eleventh, twelfth, thirteenth, and second, respectively. By inspecting the descriptors DQI01, DQI09, and DQI10, we

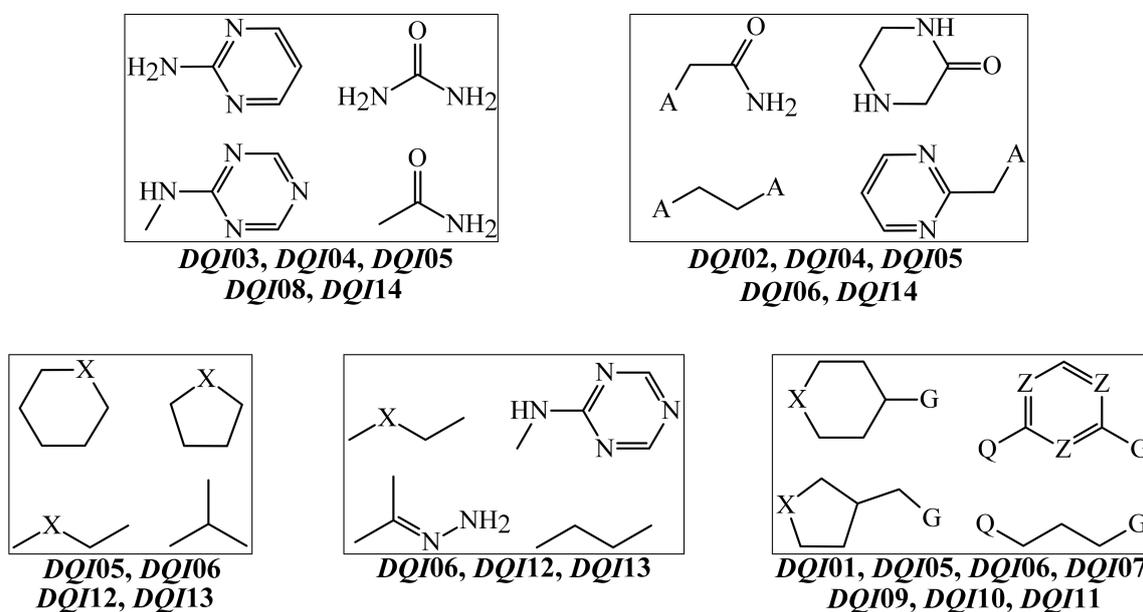
can see that they express the decrease of the joint hydrophobic contribution of any two atoms (with at least one of them being a halogen) separated by a topological distance of four, one, and two, respectively. Considering that all the halogens have a positive value of hydrophobic contribution [91], the atoms surrounding them must have a negative value of the hydrophobic contribution. Such atoms can be mainly primary and secondary carbons. Therefore, fragments such as 4-halocyclohexyl and 4-halobutyl will greatly and favorably diminish the values of  $DQI01$ ,  $DQI09$ , and  $DQI10$ , thus enhancing the inhibitory activity against the PANC-related proteins and the PANC cell lines. At the same time,  $DQI02$ ,  $DQI03$ , and  $DQI14$  characterize an increase in the joint hydrophobic contribution of any two atoms, with one of them being a heteroatom (particularly N, O, S, or P). However, while  $DQI02$  and  $DQI03$  consider fragments where the heteroatom is at the topological distance of three and four (respectively) or less from any other non-hydrogen atom,  $DQI14$  exclusively depends on the presence and number of heteroatoms. In any case, fragments containing the ethylenediamine or glycine moieties will desirably increase the value of  $DQI02$ , while fragments such as 1,3,5-triazin-2-amine, pyrimidin-2-amine, and heteroaliphatic rings (containing a nitrogen atom or being attached to it), as well as urea and carbamate, will do the same for  $DQI03$ ; all these fragments will increase the value of  $DQI14$ . On the other hand, the descriptor  $DQI12$  indicates the increase of the joint hydrophobic contribution between a carbon atom from a methyl group and its adjacent atoms (the topological distance equal to one). This means that to increase the value of  $DQI12$ , a carbon atom from a methyl group must be attached to a secondary (non-substituted) carbon or aminic nitrogen (resulting in a secondary amine).

In *Model 2*, we also have three  $D(GTI)_{cj}$  descriptors containing electronic information. One of them is  $DQI04$  (the ninth most influential), which characterizes the augmentation of the electronegativity of any two atoms (with one of them being a heteroatom) placed at the topological distance of two. Therefore, all the fragments mentioned for the descriptors  $DQI02$  and  $DQI03$  explained above cause a beneficial increase in the value of  $DQI04$ . In the case of  $DQI05$  (the fifth most important), this describes the diminution of the polar surface area in any two adjacent heteroatoms. Because  $DQI05$  directly depends on the presence of electron-rich heteroatoms, such as nitrogen and oxygen, to decrease the value of  $DQI05$ , two electron-rich heteroatoms should not be bonded. Moreover, in terms of electronic information,  $DQI08$  expresses the increase of the polar surface area of any two heteroatoms placed at the topological distance of four or less. Therefore, all the fragments mentioned for the descriptor  $DQI03$  (except for the heteroaliphatic rings) and in a lower degree, for  $DQI02$ , will increase the value of  $DQI08$ . Other groups containing sulfur (sulfonamides) and phosphorus (phosphates) will also increase the value of  $DQI08$ . This descriptor is the least significant in *Model 2*.

Steric factors play an essential role in enhancing the inhibitory activity of a chemical against the PANC-related proteins and the PANC cell lines. The  $D(GTI)_{cj}$  descriptors accounting for the steric information are  $DQI06$ ,  $DQI07$ ,  $DQI11$ , and  $DQI13$ , and, regarding their influence in *Model 2*, they ranked third, seventh, tenth, and sixth, respectively. The descriptor  $DQI06$  indicates the augmentation of the molecular size, mainly by increasing the number of aliphatic carbons but also by increasing the number of atoms (different from carbon) bonded to the aliphatic carbons. On the other hand,  $DQI07$  means that the number of halogens must be decreased; if a halogen is present, iodine, bromine, and chlorine will be preferred over fluorine. In the case of  $DQI11$ , this characterizes the decrease of the atomic weight of any two atoms (with one of them being a halogen) placed at the topological distance of two. This means that to decrease the value of  $DQI11$ , the number of halogens should be reduced, and their chemical environments must be almost exclusively constituted by carbon atoms; halogens should be attached to aliphatic or aromatic rings or in the form of 2-haloethyl moieties. For  $DQI11$ , fluorine is strongly preferred over the other halogens. Last, we have  $DQI13$ , and its value can be increased by bonding carbons from methyl groups to low-polarizability atoms, such as nitrogen and oxygen. Therefore, the presence of methoxy and methylamino groups favor the increment of  $DQI13$ .

### 3.3. Designing Multi-Protein and Multi-Cell Inhibitors as Anti-PANC Agents

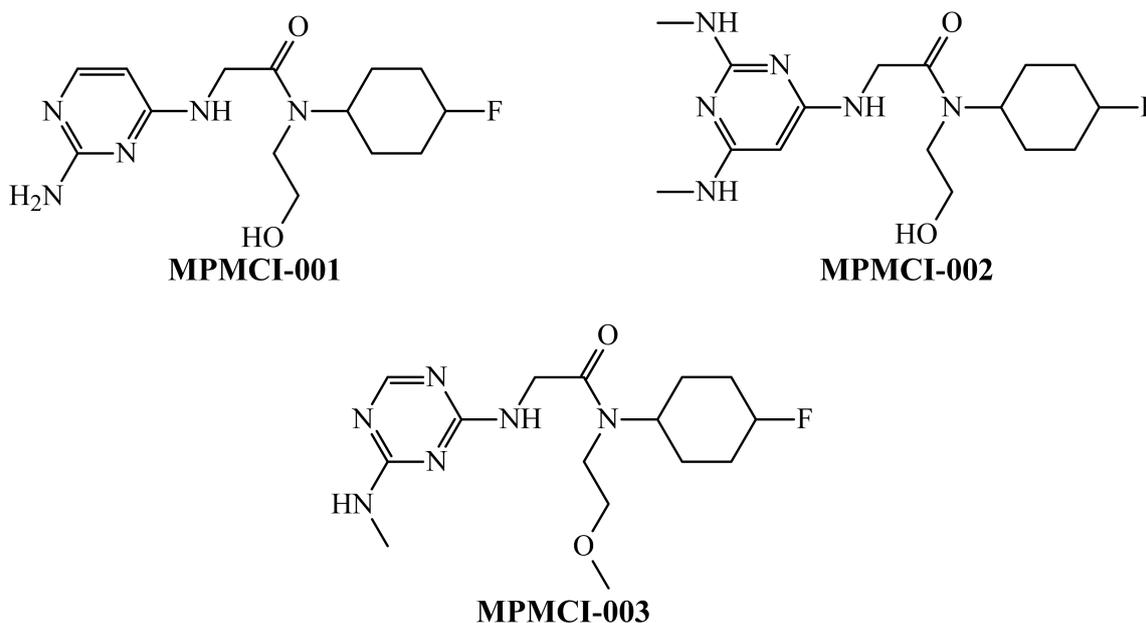
The guidelines applied to the design of new molecules from PTML models have been reported before [25,26,28,67,80,92]. Here, we used *Model 2* as a tool to design new molecules from the fragments present in Figure 3.



**Figure 3.** Generic molecular fragments directly extracted from the physicochemical and structural interpretation of the descriptors in *Model 2*. The descriptors are associated with different fragments. The symbols mean A = amino, hydroxyl, alkylamino, or alkoxy; G = halogen; Q = amino, hydroxyl, alkylamino, alkoxy, or a non-substituted secondary carbon; X = O, -NH-, or a secondary carbon; Z = N or aromatic carbon.

There were two reasons for which we preferred *Model 2* over *Model 1*. From one side, as mentioned in Section 3.1, *Model 2*, with one less descriptor, achieved a statistical performance comparable to that of *Model 1*. On the other hand, from a physicochemical point of view, *Model 2* offers a simpler (but more detailed) explanation than *Model 1*. The joint interpretation of the *D(GTI)cj* descriptors in *Model 2* suggests the presence of three well-defined regions that seem to be essential when designing a molecule. First, there should be a hydrophilic region containing several heteroatoms that can interact through hydrogen bonds. The second region, situated in the center of the molecule, may also contain heteroatoms but these should be more dispersed, with any two heteroatoms separated at a distance of three bonds. In this region, aliphatic atoms may serve as “bridges” between any two heteroatoms. The third region is expected to be hydrophobic, particularly containing a 4-halophenyl or 4-halocyclohexyl moiety, which means that the halogen will be in the periphery of the molecule.

By combining the aforementioned joint interpretation with the inspection of the fragments in Figure 3, we designed three structurally related molecules, which are depicted in Figure 4.



**Figure 4.** New molecules designed from suitable molecular fragments by using the physicochemical and structural interpretations as guidelines.

The three designed molecules were predicted by *Model 1* and *Model 2* to be anti-PANC agents through the inhibition of the PANC-related proteins and PANC cell lines. However, there were differences in the predicted probabilities. For instance, for all the 44 experimental conditions  $c_j$  reported in this work, MPMCI-001 and MPMCI-003 were predicted by both PTML-MLP models with probability values close to 100%. However, in the case of MPMCI-002, this happened only with *Model 2*. That does not mean that *Model 1* poorly predicted MPMCI-002. Notice that *Model 1* predicted MPMCI-002 to be an anti-PANC agent in 42 of the 44 experimental conditions, but the predicted probability values were lower when compared to those obtained for MPMCI-001 and MPMCI-003. This means that, in the structure of MPMCI-002, the methyl group from the methylamino moiety located between the two nitrogen atoms of the pyrimidine ring should be removed. The difference in the predicted probabilities also demonstrates the sensibility of the  $D(GTI)c_j$  descriptors in *Model 1*, allowing this model to be used as a filter of *Model 2* to ensure the correct design of new molecules. In any case, the fact that the three designed molecules were predicted by both PTML-MLP models as multi-protein and multi-cell inhibitors, theoretically confirms the aforementioned molecules to be anti-PANC agents. Moreover, the three molecules were within the applicability domain of these models. All the results of the predictions (including the assessment of the applicability domain) performed by *Model 1* and *Model 2* for the design molecules can be found in Supplementary Materials S6 and S7, respectively.

We searched for our designed molecules in different databases, such as ChEMBL [41,93], ZINC [94], and eMolecules [95], to check their novelty. In this sense, we used the similarity cutoff of 80%. No molecules similar to those designed by us were found. We would like to add that we also calculated several global physicochemical properties for the designed molecules (Table 6), which allowed us to have an idea regarding their drug-likeness.

**Table 6.** Physicochemical properties estimated for the designed molecules.

ID <sup>a</sup>	nHDon	nHAcc	MW (Da)	MlogP	AlogP	MR (cm <sup>3</sup> /mol)	nAT	RBN	PSA (Å)
MPMCI-001	4	8	311.41	1.429	0.675	83.009	44	6	104.37
MPMCI-002	4	9	354.49	2.259	1.578	99.009	52	8	102.41
MPMCI-003	2	9	340.46	2.151	1.515	92.717	49	8	92.27

<sup>a</sup> The physicochemical properties described in this table are as follows: number of hydrogen bond donors (nHDon), number of hydrogen bond acceptors (nHAcc), molecular weight (MW), logarithm of the octanol/water partition coefficient according to the Moriguchi approach (MlogP), logarithm of the octanol/water partition coefficient according to the Ghose–Crippen approach (AlogP), molar refractivity according to the Ghose–Crippen approach (MR), total number of atoms (nAT), number of rotatable bonds (RBN), and polar surface area (PSA).

In doing so, the calculated physicochemical properties allowed the use of three approaches known as the Lipinski rule of five [96], Ghose filter [97], and Veber’s guidelines [98]. We compared the values of each physicochemical property for each of the designed molecules with the cutoff values/intervals established by each of the aforementioned filters. Our designed molecules complied with all these criteria.

#### 4. Conclusions

In the context of therapeutic solutions for oncology research, finding anti-PANC chemicals is one of the greatest challenges. The multi-factorial nature of PANC is a clear indicator that both experimental and computational approaches must be focused on searching for multi-target agents to more efficiently limit or eradicate PANC growth through multiple mechanisms of action. Our computational methodology, relying on two PTML-MLP models, has established the theoretical basis to enable the fragment-based design of new molecular entities that can act as multi-protein and multi-cell inhibitors against PANC. This work consolidates the usefulness of PTML modeling as a strategy that can yield potentially new anticancer therapeutics exhibiting adequate drug-like properties.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/biomedicines10020491/s1>, Supplementary Material S1: Topological indices, averages, and standard deviation values for *Model 1*, Supplementary Material S2: Atom-based local stochastic quadratic indices, averages, and standard deviation values for *Model 2*, Supplementary Material S3: *D(GTI)cj* descriptors, classification results, and local metrics for *Model 1*, Supplementary Material S4: *D(GTI)cj* descriptors, classification results, and local metrics for *Model 2*, Supplementary Material S5: Applicability domain of the two PTML-MLP models by considering the cases/chemicals belonging to both the training and test sets, Supplementary Material S6: Topological indices, *D(GTI)cj* descriptors, classification results, and applicability domain of the designed molecules according to *Model 1*, Supplementary Material S7: Atom-based local stochastic quadratic indices, *D(GTI)cj* descriptors, classification results, and applicability domain of the designed molecules according to *Model 2*.

**Author Contributions:** Conceptualization, A.S.-P.; methodology, A.S.-P.; software, A.S.-P. and V.V.K.; validation, A.S.-P.; formal analysis, A.S.-P.; investigation, A.S.-P. and V.V.K.; resources, A.S.-P. and V.V.K.; data curation, V.V.K.; calculations, A.S.-P. and V.V.K.; writing—original draft preparation, A.S.-P. and V.V.K.; writing—review and editing, A.S.-P.; visualization, A.S.-P. and V.V.K.; supervision, A.S.-P.; project administration, A.S.-P. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All the chemical and biological (raw) data were retrieved from the public repository known as ChEMBL (<https://www.ebi.ac.uk/chembl/>, accessed on 4 January 2022).

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

## References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)] [[PubMed](#)]
2. Chang, J.S.; Chen, L.T.; Shan, Y.S.; Chu, P.Y.; Tsai, C.R.; Tsai, H.J. The incidence and survival of pancreatic cancer by histology, including rare subtypes: A nation-wide cancer registry-based study from Taiwan. *Cancer Med.* **2018**, *7*, 5775–5788. [[CrossRef](#)]
3. Klein, A.P. Pancreatic cancer epidemiology: Understanding the role of lifestyle and inherited risk factors. *Nat. Rev. Gastroenterol. Hepatol.* **2021**, *18*, 493–502. [[CrossRef](#)]
4. Cintas, C.; Douche, T.; Therville, N.; Arcucci, S.; Ramos-Delgado, F.; Basset, C.; Thibault, B.; Guillermet-Guibert, J. Signal-Targeted Therapies and Resistance Mechanisms in Pancreatic Cancer: Future Developments Reside in Proteomics. *Cancers* **2018**, *10*, 174. [[CrossRef](#)] [[PubMed](#)]
5. Hu, C.; LaDuca, H.; Shimelis, H.; Polley, E.C.; Lilyquist, J.; Hart, S.N.; Na, J.; Thomas, A.; Lee, K.Y.; Davis, B.T.; et al. Multigene Hereditary Cancer Panels Reveal High-Risk Pancreatic Cancer Susceptibility Genes. *JCO Precis. Oncol.* **2018**, *2*, 1–28. [[CrossRef](#)]
6. Dinarello, C.A. Interleukin 1 and interleukin 18 as mediators of inflammation and the aging process. *Am. J. Clin. Nutr.* **2006**, *83*, S447–S455. [[CrossRef](#)] [[PubMed](#)]
7. Raupach, B.; Peuschel, S.K.; Monack, D.M.; Zychlinsky, A. Caspase-1-mediated activation of interleukin-1beta (IL-1beta) and IL-18 contributes to innate immune defenses against *Salmonella enterica* serovar Typhimurium infection. *Infect. Immun.* **2006**, *74*, 4922–4926. [[CrossRef](#)]
8. Yang, Z.; Cao, J.; Yu, C.; Yang, Q.; Zhang, Y.; Han, L. Caspase-1 mediated interleukin-18 activation in neutrophils promotes the activity of rheumatoid arthritis in a NLRP3 inflammasome independent manner. *Jt. Bone Spine* **2016**, *83*, 282–289. [[CrossRef](#)] [[PubMed](#)]
9. Du, T.; Gao, J.; Li, P.; Wang, Y.; Qi, Q.; Liu, X.; Li, J.; Wang, C.; Du, L. Pyroptosis, metabolism, and tumor immune microenvironment. *Clin. Transl. Med.* **2021**, *11*, e492. [[CrossRef](#)]
10. Xia, S.; Zhang, Z.; Magupalli, V.G.; Pablo, J.L.; Dong, Y.; Vora, S.M.; Wang, L.; Fu, T.M.; Jacobson, M.P.; Greka, A.; et al. Gasdermin D pore structure reveals preferential release of mature interleukin-1. *Nature* **2021**, *593*, 607–611. [[CrossRef](#)]
11. Furuoka, M.; Ozaki, K.; Sadatomi, D.; Mamiya, S.; Yonezawa, T.; Tanimura, S.; Takeda, K. TNF-alpha Induces Caspase-1 Activation Independently of Simultaneously Induced NLRP3 in 3T3-L1 Cells. *J. Cell. Physiol.* **2016**, *231*, 2761–2767. [[CrossRef](#)] [[PubMed](#)]
12. Xu, S.; Li, X.; Liu, Y.; Xia, Y.; Chang, R.; Zhang, C. Inflammasome inhibitors: Promising therapeutic approaches against cancer. *J. Hematol. Oncol.* **2019**, *12*, 64. [[CrossRef](#)]
13. Daley, D.; Mani, V.R.; Mohan, N.; Akkad, N.; Pandian, G.; Savadkar, S.; Lee, K.B.; Torres-Hernandez, A.; Aykut, B.; Diskin, B.; et al. NLRP3 signaling drives macrophage-induced adaptive immune suppression in pancreatic carcinoma. *J. Exp. Med.* **2017**, *214*, 1711–1724. [[CrossRef](#)] [[PubMed](#)]
14. Schlosser, S.; Gansauge, F.; Ramadani, M.; Beger, H.G.; Gansauge, S. Inhibition of caspase-1 induces cell death in pancreatic carcinoma cells and potentially modulates expression levels of bcl-2 family proteins. *FEBS Lett.* **2001**, *491*, 104–108. [[CrossRef](#)]
15. Zhao, X.; Fan, W.; Xu, Z.; Chen, H.; He, Y.; Yang, G.; Yang, G.; Hu, H.; Tang, S.; Wang, P.; et al. Inhibiting tumor necrosis factor-alpha diminishes desmoplasia and inflammation to overcome chemoresistance in pancreatic ductal adenocarcinoma. *Oncotarget* **2016**, *7*, 81110–81122. [[CrossRef](#)]
16. Polireddy, K.; Chen, Q. Cancer of the Pancreas: Molecular Pathways and Current Advancement in Treatment. *J. Cancer* **2016**, *7*, 1497–1514. [[CrossRef](#)]
17. Subramani, R.; Lopez-Valdez, R.; Arumugam, A.; Nandy, S.; Boopalan, T.; Lakshmanaswamy, R. Targeting insulin-like growth factor 1 receptor inhibits pancreatic cancer growth and metastasis. *PLoS ONE* **2014**, *9*, e97016. [[CrossRef](#)] [[PubMed](#)]
18. Lai, E.; Puzzone, M.; Ziranu, P.; Pretta, A.; Impera, V.; Mariani, S.; Liscia, N.; Soro, P.; Musio, F.; Persano, M.; et al. New therapeutic targets in pancreatic cancer. *Cancer Treat. Rev.* **2019**, *81*, 101926. [[CrossRef](#)]
19. Ortega-Tenezaca, B.; Quevedo-Tumaili, V.; Bediaga, H.; Collados, J.; Arrasate, S.; Madariaga, G.; Munteanu, C.R.; Cordeiro, M.; Gonzalez-Diaz, H. PTML Multi-Label Algorithms: Models, Software, and Applications. *Curr. Top. Med. Chem.* **2020**, *20*, 2326–2337. [[CrossRef](#)]
20. Speck-Planche, A.; Cordeiro, M.N.D.S. Multitasking models for quantitative structure-biological effect relationships: Current status and future perspectives to speed up drug discovery. *Expert Opin. Drug Discov.* **2015**, *10*, 245–256. [[CrossRef](#)]
21. Kleandrova, V.V.; Speck-Planche, A. The urgent need for pan-antiviral agents: From multitarget discovery to multiscale design. *Future Med. Chem.* **2021**, *13*, 5–8. [[CrossRef](#)] [[PubMed](#)]
22. Halder, A.K.; Delgado, A.H.S.; Cordeiro, M. First multi-target QSAR model for predicting the cytotoxicity of acrylic acid-based dental monomers. *Dent. Mater.* **2022**, *38*, 333–346. [[CrossRef](#)]
23. Halder, A.K.; Cordeiro, M. Multi-Target in Silico Prediction of Inhibitors for Mitogen-Activated Protein Kinase-Interacting Kinases. *Biomolecules* **2021**, *11*, 1670. [[CrossRef](#)] [[PubMed](#)]
24. Bediaga, H.; Arrasate, S.; Gonzalez-Diaz, H. PTML Combinatorial Model of ChEMBL Compounds Assays for Multiple Types of Cancer. *ACS Comb. Sci.* **2018**, *20*, 621–632. [[CrossRef](#)] [[PubMed](#)]
25. Kleandrova, V.V.; Scotti, M.T.; Scotti, L.; Speck-Planche, A. Multi-Target Drug Discovery Via PTML Modeling: Applications to the Design of Virtual Dual Inhibitors of CDK4 and HER2. *Curr. Top. Med. Chem.* **2021**, *21*, 661–675. [[CrossRef](#)] [[PubMed](#)]
26. Speck-Planche, A. Multicellular Target QSAR Model for Simultaneous Prediction and Design of Anti-Pancreatic Cancer Agents. *ACS Omega* **2019**, *4*, 3122–3132. [[CrossRef](#)]

27. Diez-Alarcia, R.; Yanez-Perez, V.; Muneta-Arrate, I.; Arrasate, S.; Lete, E.; Meana, J.J.; Gonzalez-Diaz, H. Big Data Challenges Targeting Proteins in GPCR Signaling Pathways; Combining PTML-ChEMBL Models and [(35)S] GTPgammaS Binding Assays. *ACS Chem. Neurosci.* **2019**, *10*, 4476–4491. [[CrossRef](#)]
28. Kleandrova, V.V.; Speck-Planche, A. PTML Modeling for Alzheimer’s Disease: Design and Prediction of Virtual Multi-Target Inhibitors of GSK3B, HDAC1, and HDAC6. *Curr. Top. Med. Chem.* **2020**, *20*, 1657–1672. [[CrossRef](#)]
29. Ferreira da Costa, J.; Silva, D.; Caamano, O.; Brea, J.M.; Loza, M.I.; Munteanu, C.R.; Pazos, A.; Garcia-Mera, X.; Gonzalez-Diaz, H. Perturbation Theory/Machine Learning Model of ChEMBL Data for Dopamine Targets: Docking, Synthesis, and Assay of New l-Prolyl-l-leucyl-glycinamide Peptidomimetics. *ACS Chem. Neurosci.* **2018**, *9*, 2572–2587. [[CrossRef](#)]
30. Abeijon, P.; Garcia-Mera, X.; Caamano, O.; Yanez, M.; Lopez-Castro, E.; Romero-Duran, F.J.; Gonzalez-Diaz, H. Multi-Target Mining of Alzheimer Disease Proteome with Hansch’s QSBR-Perturbation Theory and Experimental-Theoretic Study of New Thiophene Isosters of Rasagiline. *Curr. Drug Targets* **2017**, *18*, 511–521. [[CrossRef](#)]
31. Martinez-Arzate, S.G.; Tenorio-Borroto, E.; Barbabosa Pliego, A.; Diaz-Albiter, H.M.; Vazquez-Chagoyan, J.C.; Gonzalez-Diaz, H. PTML Model for Proteome Mining of B-Cell Epitopes and Theoretical-Experimental Study of Bm86 Protein Sequences from Colima, Mexico. *J. Proteome Res.* **2017**, *16*, 4093–4103. [[CrossRef](#)]
32. Tenorio-Borroto, E.; Penuelas-Rivas, C.G.; Vasquez-Chagoyan, J.C.; Castanedo, N.; Prado-Prado, F.J.; Garcia-Mera, X.; Gonzalez-Diaz, H. Model for high-throughput screening of drug immunotoxicity—Study of the anti-microbial G1 over peritoneal macrophages using flow cytometry. *Eur. J. Med. Chem.* **2014**, *72*, 206–220. [[CrossRef](#)] [[PubMed](#)]
33. Vasquez-Dominguez, E.; Armijos-Jaramillo, V.D.; Tejera, E.; Gonzalez-Diaz, H. Multioutput Perturbation-Theory Machine Learning (PTML) Model of ChEMBL Data for Antiretroviral Compounds. *Mol. Pharm.* **2019**, *16*, 4200–4212. [[CrossRef](#)] [[PubMed](#)]
34. Nocado-Mena, D.; Cornelio, C.; Camacho-Corona, M.D.R.; Garza-Gonzalez, E.; Waksman de Torres, N.; Arrasate, S.; Sotomayor, N.; Lete, E.; Gonzalez-Diaz, H. Modeling Antibacterial Activity with Machine Learning and Fusion of Chemical Structure Information with Microorganism Metabolic Networks. *J. Chem. Inf. Model.* **2019**, *59*, 1109–1120. [[CrossRef](#)] [[PubMed](#)]
35. Speck-Planche, A.; Kleandrova, V.V.; Ruso, J.M.; Cordeiro, M.N.D.S. First multitarget chemo-bioinformatic model to enable the discovery of antibacterial peptides against multiple Gram-positive pathogens. *J. Chem. Inf. Model.* **2016**, *56*, 588–598. [[CrossRef](#)]
36. Speck-Planche, A.; Cordeiro, M.N.D.S. Chemoinformatics for medicinal chemistry: In silico model to enable the discovery of potent and safer anti-cocci agents. *Future Med. Chem.* **2014**, *6*, 2013–2028. [[CrossRef](#)]
37. Kleandrova, V.V.; Scotti, M.T.; Speck-Planche, A. Computational Drug Repurposing for Antituberculosis Therapy: Discovery of Multi-Strain Inhibitors. *Antibiotics* **2021**, *10*, 1005. [[CrossRef](#)]
38. Herrera-Ibata, D.M.; Pazos, A.; Orbeago-Medina, R.A.; Romero-Duran, F.J.; Gonzalez-Diaz, H. Mapping chemical structure-activity information of HAART-drug cocktails over complex networks of AIDS epidemiology and socioeconomic data of U.S. counties. *Biosystems* **2015**, *132*, 20–34. [[CrossRef](#)]
39. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Gonzalez-Diaz, H. Predicting coated-nanoparticle drug release systems with perturbation-theory machine learning (PTML) models. *Nanoscale* **2020**, *12*, 13471–13483. [[CrossRef](#)]
40. Speck-Planche, A.; Kleandrova, V.V. Chapter 16 Demystifying Artificial Neural Networks as Generators of New Chemical Knowledge: Antimalarial Drug Discovery as a Case Study. In *Machine Learning in Chemistry: The Impact of Artificial Intelligence*; The Royal Society of Chemistry: London, UK, 2020; pp. 398–423.
41. 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](#)]
42. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*, D955–D961. [[CrossRef](#)]
43. Attene-Ramos, M.S.; Austin, C.P.; Xia, M. High Throughput Screening. In *Encyclopedia of Toxicology (Third Edition)*; Wexler, P., Ed.; Academic Press Elsevier: Cambridge, MA, USA, 2014; pp. 916–917.
44. Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 1. Definition and applications for the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 844–849. [[CrossRef](#)]
45. Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 2. Molecules containing heteroatoms and QSAR applications. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 320–328. [[CrossRef](#)]
46. Estrada, E. Spectral moments of the edge adjacency matrix in molecular graphs. 3. Molecules containing cycles. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 23–27. [[CrossRef](#)]
47. Randić, M. Characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615. [[CrossRef](#)]
48. Kier, L.B.; Hall, L.H. *Molecular Connectivity in Structure-Activity Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 1986.
49. Kier, L.B.; Hall, L.H. Derivation and significance of valence molecular connectivity. *J. Pharm. Sci.* **1981**, *70*, 583–589. [[CrossRef](#)] [[PubMed](#)]
50. Estrada, E. Edge adjacency relationship and a novel topological index related to molecular volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31–33. [[CrossRef](#)]
51. Estrada, E.; Gutiérrez, Y. *MODESLAB, v1.5*; Estrada’s Laboratory: Santiago de Compostela, Spain, 2002–2004.
52. Garcia, I.; Fall, Y.; Gomez, G. Using topological indices to predict anti-Alzheimer and anti-parasitic GSK-3 inhibitors by multi-target QSAR in silico screening. *Molecules* **2010**, *15*, 5408–5422. [[CrossRef](#)]

53. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 33. [[CrossRef](#)]
54. Medina Marrero, R.; Marrero-Ponce, Y.; Barigye, S.J.; Echeverria Diaz, Y.; Acevedo-Barrios, R.; Casanola-Martin, G.M.; Garcia Bernal, M.; Torrens, F.; Perez-Gimenez, F. QuBiLS-MAS method in early drug discovery and rational drug identification of antifungal agents. *SAR QSAR Environ. Res.* **2015**, *26*, 943–958. [[CrossRef](#)]
55. Valdes-Martini, J.R.; Marrero-Ponce, Y.; Garcia-Jacas, C.R.; Martinez-Mayorga, K.; Barigye, S.J.; Vaz d'Almeida, Y.S.; Pham-The, H.; Perez-Gimenez, F.; Morell, C.A. QuBiLS-MAS, open source multi-platform software for atom- and bond-based topological (2D) and chiral (2.5D) algebraic molecular descriptors computations. *J. Cheminformatics* **2017**, *9*, 35. [[CrossRef](#)] [[PubMed](#)]
56. Marrero-Ponce, Y.; Siverio-Mota, D.; Galvez-Llompart, M.; Recio, M.C.; Giner, R.M.; Garcia-Domenech, R.; Torrens, F.; Aran, V.J.; Cordero-Maldonado, M.L.; Esguera, C.V.; et al. Discovery of novel anti-inflammatory drug-like compounds by aligning in silico and in vivo screening: The nitroindazolinone chemotype. *Eur. J. Med. Chem.* **2011**, *46*, 5736–5753. [[CrossRef](#)] [[PubMed](#)]
57. Sampaio-Dias, I.E.; Rodriguez-Borges, J.E.; Yanez-Perez, V.; Arrasate, S.; Llorente, J.; Brea, J.M.; Bediaga, H.; Vina, D.; Loza, M.I.; Caamano, O.; et al. Synthesis, Pharmacological, and Biological Evaluation of 2-Furoyl-Based MIF-1 Peptidomimetics and the Development of a General-Purpose Model for Allosteric Modulators (ALLOPTML). *ACS Chem. Neurosci.* **2021**, *12*, 203–215. [[CrossRef](#)] [[PubMed](#)]
58. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva Caracuel, E.; Gonzalez-Diaz, H. PTML Model of ChEMBL Compounds Assays for Vitamin Derivatives. *ACS Comb. Sci.* **2020**, *22*, 129–141. [[CrossRef](#)] [[PubMed](#)]
59. Tenorio-Borroto, E.; Castanedo, N.; Garcia-Mera, X.; Rivadeneira, K.; Vazquez Chagoyan, J.C.; Barbabosa Pliego, A.; Munteanu, C.R.; Gonzalez-Diaz, H. Perturbation Theory Machine Learning Modeling of Immunotoxicity for Drugs Targeting Inflammatory Cytokines and Study of the Antimicrobial G1 Using Cytometric Bead Arrays. *Chem. Res. Toxicol.* **2019**, *32*, 1811–1823. [[CrossRef](#)]
60. Ortega-Tenezaca, B.; Gonzalez-Diaz, H. IFPTML mapping of nanoparticle antibacterial activity vs. pathogen metabolic networks. *Nanoscale* **2021**, *13*, 1318–1330. [[CrossRef](#)]
61. Urista, D.V.; Carrue, D.B.; Otero, I.; Arrasate, S.; Quevedo-Tumaili, V.F.; Gestal, M.; Gonzalez-Diaz, H.; Munteanu, C.R. Prediction of Antimalarial Drug-Decorated Nanoparticle Delivery Systems with Random Forest Models. *Biology* **2020**, *9*, 198. [[CrossRef](#)]
62. Romero-Duran, F.J.; Alonso, N.; Yanez, M.; Caamano, O.; Garcia-Mera, X.; Gonzalez-Diaz, H. Brain-inspired cheminformatics of drug-target brain interactome, synthesis, and assay of TVP1022 derivatives. *Neuropharmacology* **2016**, *103*, 270–278. [[CrossRef](#)]
63. Speck-Planche, A.; Kleandrova, V.V.; Scotti, M.T. In Silico Drug Repurposing for Anti-Inflammatory Therapy: Virtual Search for Dual Inhibitors of Caspase-1 and TNF-Alpha. *Biomolecules* **2021**, *11*, 1832. [[CrossRef](#)]
64. Urias, R.W.; Barigye, S.J.; Marrero-Ponce, Y.; Garcia-Jacas, C.R.; Valdes-Martini, J.R.; Perez-Gimenez, F. IMMAN: Free software for information theory-based chemometric analysis. *Mol. Divers.* **2015**, *19*, 305–319. [[CrossRef](#)]
65. Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1895**, *58*, 240–242. [[CrossRef](#)]
66. TIBCO-Software-Inc. *STATISTICA (Data Analysis Software System)*, v13.5.0.17; TIBCO-Software-Inc: Palo Alto, CA, USA, 2018.
67. Kleandrova, V.V.; Scotti, L.; Bezerra Mendonça Junior, F.J.; Muratov, E.; Scotti, M.T.; Speck-Planche, A. QSAR Modeling for Multi-Target Drug Discovery: Designing Simultaneous Inhibitors of Proteins in Diverse Pathogenic Parasites. *Front. Chem.* **2021**, *9*, 634663. [[CrossRef](#)]
68. Speck-Planche, A.; Scotti, M.T. BET bromodomain inhibitors: Fragment-based in silico design using multi-target QSAR models. *Mol. Divers.* **2019**, *23*, 555–572. [[CrossRef](#)] [[PubMed](#)]
69. Speck-Planche, A.; Cordeiro, M.N.D.S. Speeding up Early Drug Discovery in Antiviral Research: A Fragment-Based in Silico Approach for the Design of Virtual Anti-Hepatitis C Leads. *ACS Comb. Sci.* **2017**, *19*, 501–512. [[CrossRef](#)] [[PubMed](#)]
70. Wassermann, A.M.; Nisius, B.; Vogt, M.; Bajorath, J. Identification of descriptors capturing compound class-specific features by mutual information analysis. *J. Chem. Inf. Model.* **2010**, *50*, 1935–1940. [[CrossRef](#)] [[PubMed](#)]
71. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
72. Jeffreys, H. An invariant form for the prior probability in estimation problems. *Proc. R. Soc. Lond. A Math. Phys. Sci.* **1946**, *186*, 453–461. [[CrossRef](#)]
73. Matthews, B.W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442–451. [[CrossRef](#)]
74. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Montemore, M.M.; Gonzalez-Diaz, H. PTML Model for Selection of Nanoparticles, Anticancer Drugs, and Vitamins in the Design of Drug-Vitamin Nanoparticle Release Systems for Cancer Cotherapy. *Mol. Pharm.* **2020**, *17*, 2612–2627. [[CrossRef](#)]
75. Cabrera-Andrade, A.; Lopez-Cortes, A.; Jaramillo-Koupermann, G.; Gonzalez-Diaz, H.; Pazos, A.; Munteanu, C.R.; Perez-Castillo, Y.; Tejera, E. A Multi-Objective Approach for Anti-Osteosarcoma Cancer Agents Discovery through Drug Repurposing. *Pharmaceuticals* **2020**, *13*, 409. [[CrossRef](#)]
76. Santana, R.; Zuluaga, R.; Ganan, P.; Arrasate, S.; Onieva, E.; Gonzalez-Diaz, H. Designing nanoparticle release systems for drug-vitamin cancer co-therapy with multiplicative perturbation-theory machine learning (PTML) models. *Nanoscale* **2019**, *11*, 21811–21823. [[CrossRef](#)]
77. Casanola-Martin, G.M.; Le-Thi-Thu, H.; Perez-Gimenez, F.; Marrero-Ponce, Y.; Merino-Sanjuan, M.; Abad, C.; Gonzalez-Diaz, H. Multi-output model with Box-Jenkins operators of linear indices to predict multi-target inhibitors of ubiquitin-proteasome pathway. *Mol. Divers.* **2015**, *19*, 347–356. [[CrossRef](#)] [[PubMed](#)]

78. Marzaro, G.; Chilin, A.; Guiotto, A.; Uriarte, E.; Brun, P.; Castagliuolo, I.; Tonus, F.; Gonzalez-Diaz, H. Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors. *Eur. J. Med. Chem.* **2011**, *46*, 2185–2192. [[CrossRef](#)] [[PubMed](#)]
79. Munteanu, C.R.; Magalhaes, A.L.; Uriarte, E.; Gonzalez-Diaz, H. Multi-target QPDR classification model for human breast and colon cancer-related proteins using star graph topological indices. *J. Theor. Biol.* **2009**, *257*, 303–311. [[CrossRef](#)]
80. Speck-Planche, A. Combining Ensemble Learning with a Fragment-Based Topological Approach to Generate New Molecular Diversity in Drug Discovery: In Silico Design of Hsp90 Inhibitors. *ACS Omega* **2018**, *3*, 14704–14716. [[CrossRef](#)] [[PubMed](#)]
81. Estrada, E. How the parts organize in the whole? A top-down view of molecular descriptors and properties for QSAR and drug design. *Mini Rev. Med. Chem.* **2008**, *8*, 213–221. [[CrossRef](#)] [[PubMed](#)]
82. Estrada, E.; Patlewicz, G.; Gutierrez, Y. From knowledge generation to knowledge archive. A general strategy using TOPS-MODE with DEREK to formulate new alerts for skin sensitization. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 688–698. [[CrossRef](#)]
83. Estrada, E.; Molina, E. Automatic extraction of structural alerts for predicting chromosome aberrations of organic compounds. *J. Mol. Graph. Model.* **2006**, *25*, 275–288. [[CrossRef](#)]
84. Ghose, A.K.; Crippen, G.M. Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21–35. [[CrossRef](#)]
85. Estrada, E. Physicochemical Interpretation of Molecular Connectivity Indices. *J. Phys. Chem. A* **2002**, *106*, 9085–9091. [[CrossRef](#)]
86. Randić, M.; Zupan, J. On interpretation of well-known topological indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 550–560. [[CrossRef](#)] [[PubMed](#)]
87. Estrada, E. Edge adjacency relationships in molecular graphs containing heteroatoms: A new topological index related to molar volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 701–707. [[CrossRef](#)]
88. Estrada, E.; Rodríguez, L. Edge-Connectivity Indices in QSPR/QSAR Studies. 1. Comparison to Other Topological Indices in QSPR Studies. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1037–1041. [[CrossRef](#)]
89. Matsson, P.; Kihlberg, J. How Big Is Too Big for Cell Permeability? *J. Med. Chem.* **2017**, *60*, 1662–1664. [[CrossRef](#)] [[PubMed](#)]
90. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E.A. Atom, atom-type, and total nonstochastic and stochastic quadratic fingerprints: A promising approach for modeling of antibacterial activity. *Bioorg. Med. Chem.* **2005**, *13*, 2881–2899. [[CrossRef](#)]
91. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102*, 3762–3772. [[CrossRef](#)]
92. Kleandrova, V.V.; Scotti, M.T.; Scotti, L.; Nayariseri, A.; Speck-Planche, A. Cell-based multi-target QSAR model for design of virtual versatile inhibitors of liver cancer cell lines. *SAR QSAR Environ. Res.* **2020**, *31*, 815–836. [[CrossRef](#)]
93. Gaulton, A.; Bellis, L.J.; Bento, A.P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; et al. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100–D1107. [[CrossRef](#)]
94. Irwin, J.J.; Shoichet, B.K. ZINC—A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. [[CrossRef](#)]
95. Gubernator, K.; James, C.A.; Gubernator, N. eMolecules. Available online: <https://www.emolecules.com/> (accessed on 20 January 2022).
96. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [[CrossRef](#)]
97. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. A knowledge-based approach in designing combinatorial or medicinal chemistry libraries for drug discovery. 1. A qualitative and quantitative characterization of known drug databases. *J. Comb. Chem.* **1999**, *1*, 55–68. [[CrossRef](#)] [[PubMed](#)]
98. Veber, D.F.; Johnson, S.R.; Cheng, H.Y.; Smith, B.R.; Ward, K.W.; Kopple, K.D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615–2623. [[CrossRef](#)] [[PubMed](#)]