ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

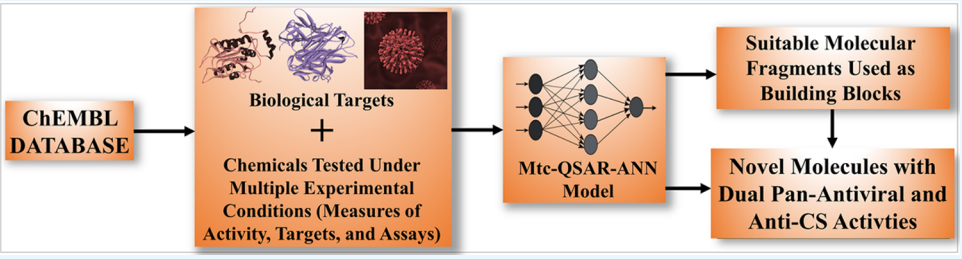
Multi-Condition QSAR Model for the Virtual Design of Chemicals with Dual Pan-Antiviral and Anti-Cytokine Storm Profiles

Alejandro Speck-Planche* and Valeria V. Kleandrova

Cite This: *ACS Omega* 2022, 7, 32119–32130

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Respiratory viruses are infectious agents, which can cause pandemics. Although nowadays the danger associated with respiratory viruses continues to be evidenced by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as the virus responsible for the current COVID-19 pandemic, other viruses such as SARS-CoV-1, the influenza A and B viruses (IAV and IBV, respectively), and the respiratory syncytial virus (RSV) can lead to globally spread viral diseases. Also, from a biological point of view, most of these viruses can cause an organ-damaging hyperinflammatory response known as the cytokine storm (CS). Computational approaches constitute an essential component of modern drug development campaigns, and therefore, they have the potential to accelerate the discovery of chemicals able to simultaneously inhibit multiple molecular and nonmolecular targets. We report here the first multicondition model based on quantitative structure−activity relationships and an artificial neural network (mtc-QSAR-ANN) for the virtual design and prediction of molecules with dual pan-antiviral and anti-CS profiles. Our mtc-QSAR-ANN model exhibited an accuracy higher than 80%. By interpreting the different descriptors present in the mtc-QSAR-ANN model, we could retrieve several molecular fragments whose assembly led to new molecules with drug-like properties and predicted pan-antiviral and anti-CS activities.

1. INTRODUCTION

Respiratory viruses represent infectious agents that can lead to life-threatening medical conditions. The danger associated with infections caused by respiratory viruses continues to be demonstrated by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1] as the virus responsible for the current COVID-19 pandemic. In any case, although SARS-CoV-1, SARS-CoV-2, and the influenza A virus (IAV) are some of the most studied viral pathogens[1−4] due to their transmissibility and lethality worldwide, evolved strains of other respiratory viruses such as the influenza B virus (IBV) and the respiratory syncytial virus (RSV) have also the potential to cause epidemics or pandemics.[5,6] Two alarming aspects associated with the infections caused by all the respiratory viruses mentioned above deserve special attention. On one side, the presence of at least two of these respiratory viruses can involve coinfection,[7−9] a poorly understood phenomenon that can lead to an increment in morbidity and mortality while also making antiviral drugs to became less effective. On the other hand, most of these viruses can cause the organ-damaging hyperinflammatory response known as the cytokine storm (CS),[10−12] where the proteins known as caspase-1 and tumor necrosis factor-alpha (TNF-alpha) play a determinant role. In this sense, caspase-1 is essential for the maturation and/or release of two CS inflammatory proteins[13−15] named interleukin 1β (IL-1β) and interleukin 18 (IL-18) while TNF-alpha is a CS protein itself,[16] triggering several inflammation-related proteins such as caspase-1[17] and the CS protein named interleukin 6 (IL-6).[18] All these elements, together with the fact that the current antiviral therapies are very narrow in terms of mechanisms of action and the number of targeted viruses, indicate the need of finding new chemicals with the ability to act as pan-antivirals[19] and to simultaneously inhibit the CS proteins caspase-1 and TNF-alpha.[20]

Nowadays, it is well-known that computational methods are of paramount importance in modern drug development.[21]

Received: May 30, 2022
Accepted: August 19, 2022
Published: August 29, 2022

ACS Publications

© 2022 The Authors. Published by American Chemical Society

Table 1. Different Experimental Conditions under Which the Molecules of the Present Dataset Were Assayed

cj[a]	ma[b]	cutoff value (nM)[c]	bt[d]	ai[e]
c1	EC_{50} (nM)	≤3800	IAV (A-Puerto Rico-8-1934 (H1N1))	F (organism-based format)
c2	EC_{50} (nM)		IAV (A-Puerto Rico-8-1934 (H1N1))	F (cell-based format)
c3	EC_{50} (nM)		IAV (A-Puerto Rico-8-1934 (H1N1))	F (assay format)
c4	EC_{50} (nM)	≤10000	IBV	F (organism-based format)
c5	EC_{50} (nM)		IBV	F (cell-based format)
c6	EC_{50} (nM)	≤300	RSV	F (organism-based format)
c7	EC_{50} (nM)		RSV	F (cell-based format)
c8	EC_{50} (nM)	≤7500	SARS-CoV-1	F (organism-based format)
c9	EC_{50} (nM)	≤1200	SARS-CoV-2	F (organism-based format)
c10	IC_{50} (nM)	≤8600	IAV (A-Puerto Rico-8-1934 (H1N1))	F (organism-based format)
c11	IC_{50} (nM)		IAV (A-Puerto Rico-8-1934 (H1N1))	F (cell-based format)
c12	IC_{50} (nM)	≤10000	IBV	F (organism-based format)
c13	IC_{50} (nM)	≤3600	RSV	F (organism-based format)
c14	IC_{50} (nM)	≤1080	SARS-CoV-1	F (organism-based format)
c15	IC_{50} (nM)	≤6070	SARS-CoV-2	F (organism-based format)
c16	IC_{50} (nM)p	≤1100	Caspase-1	B (assay format)
c17	IC_{50} (nM)p		Caspase-1	B (single protein format)
c18	IC_{50} (nM)p		Caspase-1	B (cell-based format)
c19	IC_{50} (nM)p	≤1635	TNF-alpha	B (single protein format)
c20	IC_{50} (nM)p		TNF-alpha	F (assay format)
c21	IC_{50} (nM)p		TNF-alpha	B (assay format)
c22	IC_{50} (nM)p		TNF-alpha	B (cell-based format)
c23	IC_{50} (nM)p		TNF-alpha	F (cell-based format)

[a]Codes for the different experimental conditions *cj*, which are combinations of the elements *ma* (measures of activity), *bt* (biological targets), and *ai* (assay information containing diverse experimental protocols). [b]Measures of inhibitory activity. EC_{50} (nM) is the effective concentration leading to a 50% reduction in the cytophaticity caused by a virus or inhibition of viral replication, IC_{50} (nM) is the concentration required for 50% inhibition of the virus, and IC_{50} (nM)p is the concentration required for 50% inhibition of a protein associated with the cytokine storm. [c]Value of activity from which a chemical was annotated as active [$BAi(cj) = 1$]. [d]Targets (respiratory viruses or proteins associated with the cytokine storm). [e]Assay information related to the different test protocols. Each annotation combines the columns "assay type" (first letter) and "BioAssay Ontology" (phrase between parentheses), which were extracted from the ChEMBL file containing inhibitory activity data.

However, in the context of finding new chemicals with the potential to treat viral infections, *in silico* approaches such as quantitative structure–activity relationships (QSAR), 3D-QSAR, molecular docking, molecular dynamics simulation, virtual screening, and/or shape-based pharmacophore have been applied to discover molecules with either antiviral activity[22−28] or inhibitory potency against CS-related proteins,[29−32] but never both biological profiles at the same time. Also, such computational approaches have at least one or more limitations, such as the use of small data sets of compounds belonging to the same chemical family, the study of inhibitory data based on only one biological target (a virus, a viral protein, or a CS-related biomolecule), the reliance on only one assay protocol, and the lack of sufficient physicochemical and structural information, which can provide deeper insight when searching for new molecular entities with the desired bioactivities.

Because of the significance of computational drug discovery, over the last 10 years, several research groups have emphasized the development and application of the methodology known as Perturbation Theory and Machine Learning (PTML)[33] through which it has been possible to overcome all the drawbacks of the *in silico* approaches mentioned above. Thus, PTML models have been able to integrate chemical and biological data at different levels of complexity in scientific areas as diverse as antimicrobial research,[34−38] oncology,[39−42] neurosciences,[43] immunology,[44] and peptide/protein science.[45−47] In doing so, PTML models have been able to predict multiple activities, toxicities, and/or pharmacokinetic end points, while considering dissimilar biological targets (e.g., proteins, microorganisms, cells, rodents, etc.) and many assay protocols. To date, there is no computational model reported in the scientific literature able to guide the search for chemicals with both pan-antiviral and anti-CS activities. Bearing in mind all these ideas, we report here, for the first time, the theoretical foundations of the PTML methodology in the context of both pan-antiviral and anti-inflammatory therapies. Particularly, we have created a multicondition QSAR model based on artificial neural networks (mtc-QSAR-ANN) as a tool for the virtual design and prediction of drug-like molecules with dual pan-antiviral and anti-CS profiles.

2. RESULTS AND DISCUSSION

2.1. The Mtc-QSAR-ANN Model. In this work, we developed an mtc-QSAR-ANN model able to consider 23 different experimental conditions (cj) when predicting the inhibitory activity against either multiple respiratory viruses or the CS proteins caspase-1 and TNF-alpha (Table 1). Each of those experimental conditions cj is a combination of three elements: the measure of inhibitory activity (ma), the biological target on which the experiment was performed (bt), and the assay information involving different test protocols (ai). Notice that the mtc-QSAR-ANN model used a classification approach, predicting molecules with $BAi(cj) = 1$ (active) or $BAi(cj) = -1$ (inactive), where $BAi(cj)$ represented the categorical variable of inhibitory activity. All the chemical and biological data used in this work can be found in Supporting Information (Tables S1 and S2).

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

Table 2. Symbols and Codes of the $D[GTI]cj$ Descriptors Present in the mtc-QSAR-ANN Model

symbology[a]	code[b]	concept
$D[Xv(C)6]ma$	$DGTI01$	deviation of the Kier–Hall (valence) connectivity index based only on cluster subgraphs of order 6
$D[NXv(C)4]ma$	$DGTI02$	deviation of the normalized Kier–Hall (valence) connectivity index based only on cluster subgraphs of order 4
$D[Xv(P)6]bt$	$DGTI03$	deviation of the Kier–Hall (valence) connectivity index based only on path subgraphs of order 6
$D[NSM(Gas)1]bt$	$DGTI04$	deviation of the normalized spectral moment of order 1 based on bonds weighted by the Gasteiger–Marsili charges.
$D[NSM(Gas)2]bt$	$DGTI05$	deviation of the normalized spectral moment of order 2 based on bonds weighted by the Gasteiger–Marsili charges.
$D[NXv(PC)4]bt$	$DGTI06$	deviation of the normalized Kier–Hall (valence) connectivity index based only on path-cluster subgraphs of order 4
$D[Ne(P)5]bt$	$DGTI07$	deviation of the normalized edge (bond) connectivity index based only on path subgraphs of order 5
$D[NJ]bt$	$DGTI08$	deviation of the normalized Balaban index.
$D[SM(Hyd)1]ai$	$DGTI09$	deviation of the spectral moment of order 1 based on bonds weighted by the hydrophobicity contributions.
$D[3k(alpha)]ai$	$DGTI10$	deviation of Kier's shape index based only on path subgraphs of order 3
$D[J]ai$	$DGTI11$	deviation of the Balaban index.
$D[NSM(Psa)6]ai$	$DGTI12$	deviation of the normalized spectral moment of order 6 based on bonds weighted by the polar surface area.
$D[NSM(Mol)5]ai$	$DGTI13$	deviation of the normalized spectral moment of order 5 based on bonds weighted by the molar refractivity.
$D[NXv(Ch)5]ai$	$DGTI14$	deviation of the normalized Kier–Hall (valence) connectivity index based only on chain subgraphs of order 5
$D[NKFI]ai$	$DGTI15$	deviation of the normalized Kierflexibility index.

[a]The $D[GTI]cj$ descriptors with ending "*ma*" characterize both the molecular structure and the measures of inhibitory activity. Those $D[GTI]cj$ descriptors with the ending "*bt*" describe the chemical structure as well as the biological targets (respiratory viruses and proteins associated with the cytokine storm). Finally, the $D[GTI]cj$ descriptors with ending "*ai*" characterize the chemical structure and information related to different experimental assay protocols. [b]From now on, for the sake of simplicity, the codes will be used instead of the original symbols to explain either the statistical significance or the physicochemical interpretation of the $D[GTI]cj$ descriptors.

The best mtc-QSAR-ANN model found by us has the notation MLP 15-68-2, which means that our model is based on a multilayer perceptron network containing 15 nodes equivalent to the 15 multicondition descriptors of the type $D[GTI]cj$, which entered in the mtc-QSAR-ANN model (Table 2). According to such a notation, 68 neurons were used in the hidden layer containing a hyperbolic tangent function while the number two indicates that the mtc-QSAR-ANN model could predict the two aforementioned values of $BAi(cj)$ in the output layer by relying on a softmax function.

The mtc-QSAR-ANN model displayed a good global performance, with accuracy (Acc) values of 85.14% and 80.19% in training and test sets, respectively. Other statistical indices such as sensitivity $[Sn(\%)]$, specificity $[Sp(\%)]$, and Matthews' correlation coefficient (MCC)[48] supported the good statistical quality and predictive power of our mtc-QSAR-ANN model (Table 3). For instance, $[Sn(\%)]$ and $[Sp(\%)]$ were higher than 79%. Also, MCC values were relatively close to one, which indicated a strong convergence between the predicted and observed values of the categorical variable of biological activity $BAi(cj)$.

Table 3. Statistical Performance of the mtc-QSAR-ANN Model

symbols[a]	training set	test set
N_{Active}	1156	374
CCC_{Active}	981	297
Sn (%)	84.86	79.41
$N_{Inactive}$	1435	469
$CCC_{Inactive}$	1225	379
Sp (%)	85.37	80.81
MCC	0.700	0.600

[a]N_{Active}, number of molecules/cases labeled as active; $N_{Inactive}$, number of molecules/cases annotated as inactive; CCC_{Active}, number of molecules/cases correctly classified as active; $CCC_{Inactive}$, number of molecules/cases correctly classified as inactive; Sn (%), statistical sensitivity (percentage of molecules/cases correctly classified as active); Sp (%), statistical specificity (percentage of molecules/cases correctly classified as inactive); MCC, Matthews' correlation coefficient.

We also examined the local sensitivities and specificities, which depended on specific elements. In the training set, in the case of the measures of activity (ma), the sensitivity $[Sn(\%)]ma$ and the specificity $[Sp(\%)]ma$ were higher than or equal to 79.52% and 83.78%, respectively. Both magnitudes exhibited values above 72% in the test set. Considering the local measures in the case of the biological targets (bt), $[Sn(\%)]bt$ and $[Sp(\%)]bt$ had acceptable values in the range 62.86%–92% for the whole data set. The only exception was the biological target SARS-CoV-1, for which $[Sn(\%)]bt$ = 44.44% was achieved in the test set. Last, in the case of the assay information involving different experimental protocols (ai), the $[Sn(\%)]ai$ and $[Sp(\%)]ai$ values were in the interval 75%–100% in either training or test sets. Only $[Sn(\%)]ai$ for the assay information denoted as "F (assay format)" was below this range, with values of 53.49% and 38.46% for training and test sets, respectively. We suggest that the incorrectly classified/predicted chemicals are a consequence of the lack of universality of our $D[GTI]cj$ descriptors (also valid for any molecular descriptor reported to date in the scientific literature), as well as the complexity of the experimental data involving different measures of activity (ma), multiple targets (bt), and great variability of the assay information (ai). In any case, from the analyses of both global and local statistical indices, we can infer that our mtc-QSAR-ANN model has a good statistical quality and predictive power. Detailed information regarding the $D[GTI]cj$ descriptors and the classification results is available in Supporting Information (Tables S3–S6).

In addition to analyzing the performance of the mtc-QSAR-ANN model, we also assessed the reliability of its classifications/predictions by determining the applicability domain according to the descriptors' space approach.[49,50] In doing so, we defined a local score of applicability domain (LSAD) for each $D[GTI]cj$ descriptor as well as a total score (TSAD),[50,51] which was the sum of the LSAD values. Because our mtc-QSAR-ANN model contains 15 $D[GTI]cj$ descriptors,

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119–32130

it was expected that only molecules/cases with TSAD = 15 would be those falling within the applicability domain. Of the 3434 molecules/cases in the data set, only four were outside the applicability domain (TSAD = 14). Yet, the deletion of these four outliers did not improve the performance of the mtc-QSAR-ANN model (Supporting Information Table S7).

Last, we would like to emphasize that although the present mtc-QSAR-ANN model is based on a single neural network, when performing virtual screening, it can behave as a consensus tool. Notice that the previous analysis of the local measures $[Sn(\%)]ma$, $[Sp(\%)]ma$, $[Sn(\%)]bt$, $[Sp(\%)]bt$, $[Sn(\%)]ai$, and $[Sp(\%)]ai$ demonstrates that our mtc-QSAR-ANN model can predict the activity of any molecule against a defined target by considering several experimental conditions (see Table 1). Therefore, if for a given molecule, we combine the results and the reliability (the latter provided by the applicability domain) of such predictions against a specific target under different experimental conditions, it will be possible to obtain a consensus idea on the activity or inactivity of that molecule against the target under study.

2.2. Physicochemical and Structural Interpretation of the Molecular Descriptors. Before using our mtc-QSAR-ANN model as a tool to design new molecules, we interpreted the different $D[GTI]cj$ descriptors from a physicochemical and structural point of view. In doing so, for each $D[GTI]cj$ descriptor present in the mtc-QSAR-ANN model, we calculated two mean values, one for the chemicals annotated and correctly classified as active and the other for those molecules considered and accurately classified as inactive (Table 4).[50,51] Such calculations were carried out by considering only the molecules in the training set. The comparison of the two mean values permitted us to gain insight into how the value of each $D[GTI]cj$ descriptor should be varied (increased or diminished) to increase both the pan-antiviral and anti-CS activities.

Table 4. Tendencies of Variation of the $D[GTI]cj$ Descriptors in the mtc-QSAR-ANN Model

descriptors	class-based means[a]		tendency[b]
	active	inactive	
*DGTI*01	1.4391×10^{-2}	5.9700×10^{-2}	decrease
*DGTI*02	2.8440×10^{-2}	3.1588×10^{-2}	decrease
*DGTI*03	1.6610×10^{-3}	-5.2028×10^{-2}	increase
*DGTI*04	-3.1118×10^{-2}	1.2448×10^{-1}	decrease
*DGTI*05	1.0951×10^{-2}	-1.2250×10^{-1}	increase
*DGTI*06	1.1698×10^{-2}	1.0999×10^{-1}	decrease
*DGTI*07	-2.9689×10^{-2}	-3.8392×10^{-2}	increase
*DGTI*08	1.3665×10^{-2}	1.3667×10^{-1}	decrease
*DGTI*09	-2.4072×10^{-2}	8.7241×10^{-2}	decrease
*DGTI*10	4.9842×10^{-3}	-2.3010×10^{-2}	increase
*DGTI*11	1.6154×10^{-2}	4.7476×10^{-2}	decrease
*DGTI*12	1.3308×10^{-2}	-1.1988×10^{-3}	increase
*DGTI*13	1.4227×10^{-3}	9.9303×10^{-2}	decrease
*DGTI*14	8.0425×10^{-3}	2.4919×10^{-2}	decrease
*DGTI*15	8.4249×10^{-3}	4.8861×10^{-2}	decrease

[a]These are the averages calculated for each $D[GTI]cj$ descriptor by considering chemicals (from the training set) belonging to a defined class (active or inactive). [b]Variation of the value of a $D[GTI]cj$ descriptor that should be expected to increase the inhibitory activity against the respiratory viruses and the proteins associated with the cytokine storm.

We also relied on the use of a graphic containing the sensitivity values (*SVs*) of the $D[GTI]cj$ descriptors, which indicated their degrees of importance in the mtc-QSAR-ANN model (Figure 1). The $D[GTI]cj$ descriptors with the largest *SVs* are the most influential, and therefore, they represent the most important physicochemical properties and structural features that a molecule should have to enhance its pan-antiviral and anti-CS activities.

We have five $D[GTI]cj$ descriptors derived from the atom-based connectivity indices, and therefore, they characterize the molecular accessibility of the chemicals,[52,53] i.e., the ability of certain regions/fragments/functional groups of the chemicals to interact with the different biological targets (or components of them). These $D[GTI]cj$ descriptors are *DGTI*01, *DGTI*02, *DGTI*03, *DGTI*06, and *DGTI*14. Thus, to modify the molecular accessibility in the way of increasing the pan-antiviral and anti-CS activities, the following physicochemical and structural requirements are needed. First, the number of functional groups of the type G−JL_3 should be avoided. In such functional groups, G, J, and L are non-hydrogen atoms, and the three L atoms can be either equal or different. In G−JL_3, when G is attached to at least two non-hydrogen atoms besides J (e.g., the trifluoromethyl attached to a benzene ring), the information on the diminution of the molecular accessibility is provided by the decrease of the values of *DGTI*01 and *DGTI*02. If the atom G is attached to only one non-hydrogen atom besides J (the same trifluoromethyl group attached to carbon from a methylene group), then, the information will be described only by the diminution of *DGTI*02 (its value is also reduced by increasing the total number of non-hydrogen atoms in a molecule). We would like to highlight that if for some reason, the functional groups G−JL_3 are present in a molecule, then G, J, and L should be preferably noncarbon atoms such as oxygen or nitrogen. The descriptors *DGTI*01 and *DGTI*02 rank 12th and 13th among the most important $D[GTI]cj$ descriptors in the mtc-QSAR-ANN model, respectively.

In addition, we have *DGTI*03 (the ninth most important descriptor), which indicates the increase in the number of linear fragments containing six bonds (hexane-like skeletons). In such fragments, atoms such as S and P, as well as aliphatic portions and aromatic atoms, are also highly beneficial. In this sense, Figure 2 depicts a series of generic molecular fragments whose presence can positively favor the variation of the values of the $D[GTI]cj$ descriptors, including *DGTI*03.

In the case of *DGTI*06 and *DGTI*14, they measure the diminution of the molecular accessibility in regions containing methylbutane-like skeletons and five-membered rings, respectively. This means that to increase the pan-antiviral and anti-CS activities, the number of those fragments must be reduced as much as possible; in the case where they are present, those fragments should preferably contain noncarbon atoms. The values of *DGTI*06 and *DGTI*14 can also be diminished by augmenting the number of non-hydrogen atoms in a molecule. We would like to notice that *DGTI*06 and *DGTI*14 are the second and fifth most influential $D[GTI]cj$ descriptors in the mtc-QSAR-ANN model, respectively.

Our mtc-QSAR-ANN model also contains five $D[GTI]cj$ descriptors (*DGTI*04, *DGTI*05, *DGTI*09, *DGTI*12, and *DGTI*13) based on the spectral moments of the bond adjacency matrix,[54−59] which means that they characterize how much any given physicochemical property is concentrated in different regions of a molecule. From one side, the favorable

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

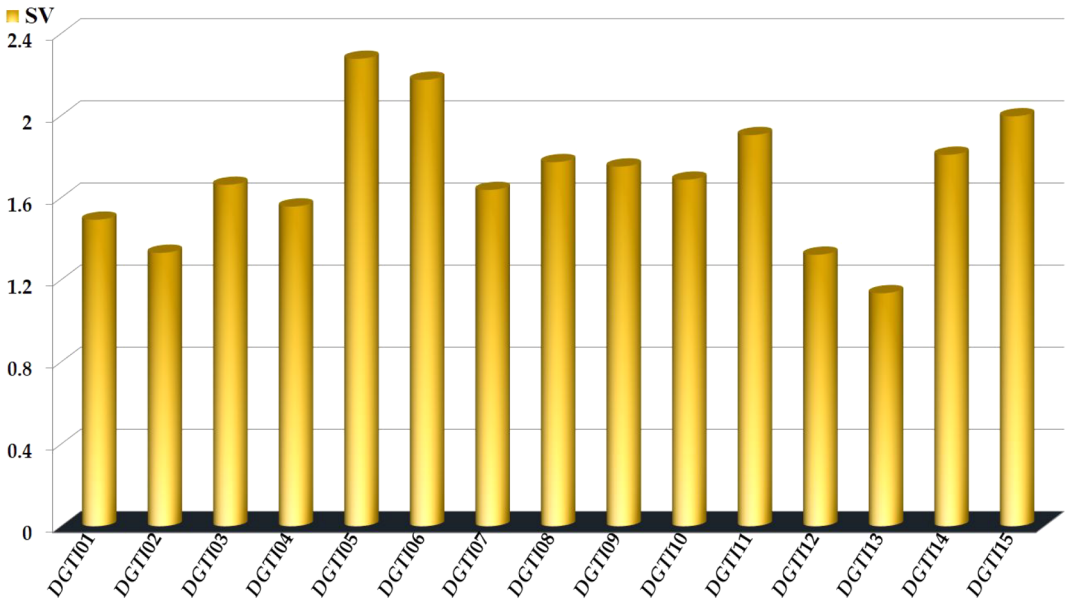


Figure 1. Sensitivity values as measures of the importance of the *D*[*GTI*]*cj* descriptors present in the mtc-QSAR-ANN model.

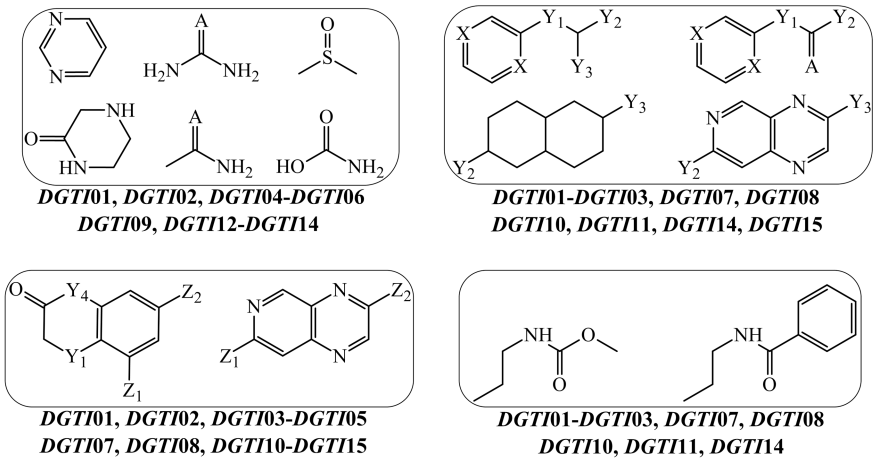


Figure 2. Different molecular fragments whose presence favorably affects the values of the *D*[*GTI*]*cj* descriptors. The symbols have the following meanings: A = O or S; X = C or N; Y_1 = O, $-CH_2-$ or $-NH-$; Y_2 and Y_3 can be any atom; Y_4 = O or $-NH-$; Z_1 and Z_2 can be F or any functional group whose electronegative atom (can be only O or N) is the one attached to the aromatic ring.

diminution of the value of *DGTI*04 (ranked 11th in terms of significance) is equivalent to increasing the number of electronegative atoms (mainly N, O, S, F, and Cl) of a molecule. The presence of these same atoms favorably increases the value of *DGTI*05 but these electronegative atoms should exist in fragments formed by two bonds (without counting bond multiplicity). We would like to highlight that *DGTI*05 is the most influential *D*[*GTI*]*cj* descriptor in the mtc-QSAR-ANN model and the presence of fragments such as amide, ester, urea, carbamate, and trifluoromethyl, heteroaromatic rings such as pyrimidine and pyridazine, and benzene rings substituted with fluorine atoms or polar groups (amine, amide, hydroxyl, urea, carbamate, etc.) can desirably increase the value of *DGTI*05. On the other hand, *DGTI*09 (the seventh most important) accounts for the global hydrophobicity of a molecule, and thus, most of the fragments favoring *DGTI*05 will also be suitable for *DGTI*09 (except for ester, trifluoromethyl, and fluorobenzene as well as alkyl groups), including benzene rings where a hydrogen atom has been replaced by nitrogen or oxygen. We also have *DGTI*12, which describes the augmentation of the polar surface area in fragments containing six bonds or less. In particular, the value of *DGTI*12 (ranking 14th in terms of significance) can be increased through the presence of any functional group containing the elements N, O, S, and/or P (e.g., amine, amide, ester, carbamate, urea, sulfonamide, sulfoxide, sulfone, and phosphate) and the aforementioned heteroaromatic rings. For the case of *DGTI*13 (being the least important descriptor), its decrease is equivalent to the presence of low-refractivity atoms such as N, O, and F. Consequently, the presence of functional groups such as amide, primary and secondary amines, and pyrido[4,3-*b*]pyrazines (with a nitrogen or oxygen atom replacing hydrogen at positions 3 and 7) is encouraged. However, because *DGTI*13 encompasses fragments having five bonds or less (with emphasis on four-bond fragments), functional groups with high-refractivity atoms (S, P, and halogens except for fluor) should be avoided, mainly, sulfonamides, sulfones, and phosphates (or any phosphorus-

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

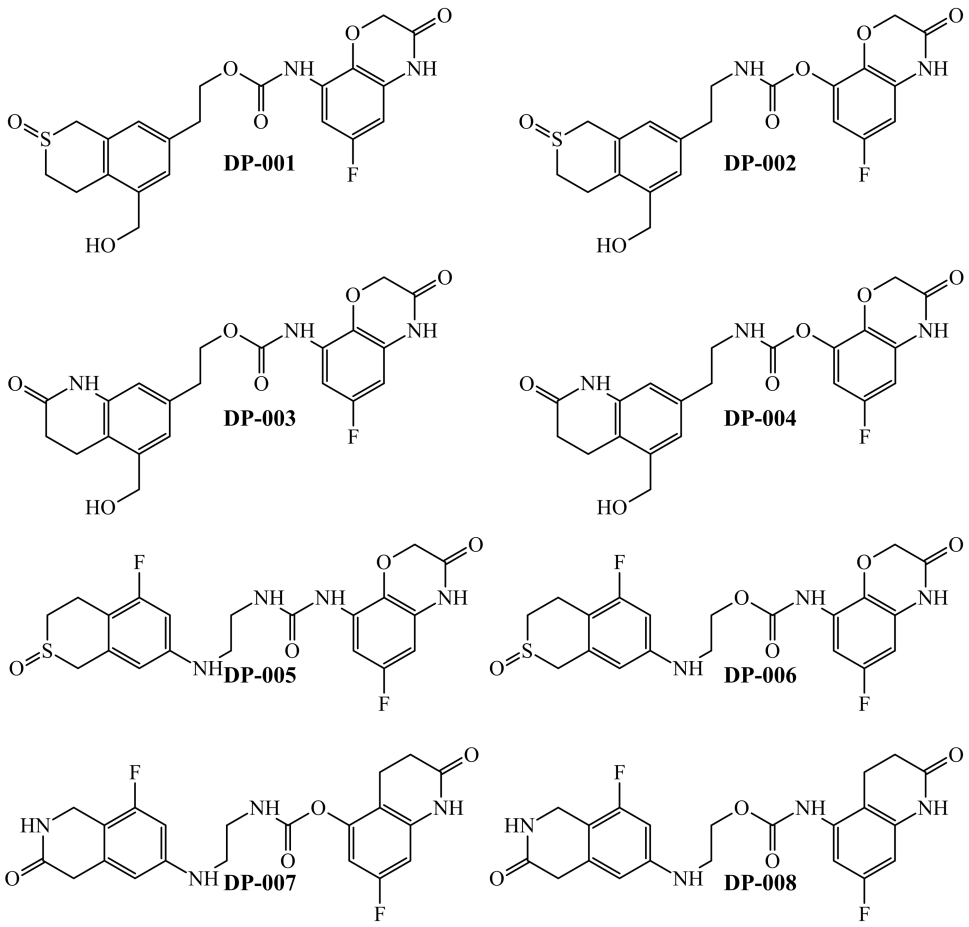


Figure 3. New molecules designed by assembling several fragments according to the physicochemical and structural interpretation of the *D*[*GTI*]*cj* descriptors.

based group where phosphorus is attached to other four non-hydrogen atoms). If a functional group containing a high-refractivity atom is present, that group should be in the periphery of a molecule.

The last five *D*[*GTI*]*cj* descriptors contain structural information which is different from the *D*[*GTI*]*cj* descriptors already discussed (see Figure 2). For instance, *DGTI*07 is based on the edge-connectivity index and, as such, is a measure of the molecular volume of a chemical.[60−62] Specifically, *DGTI*07 (ranked the tenth most influential) indicates the increment of the number of linear fragments containing five bonds (pentane-like skeletons). We have also *DGTI*08 and *DGTI*11, which are derived from the Balaban index,[63] and therefore, they are indicators of the shape of a molecule. The diminution of the values of *DGTI*08 and *DGTI*11 converges with the decrease of the number of ramifications in the molecules and the increase in the number of rings. From a structural point of view, the only difference between *DGTI*08 and *DGTI*11 is that the former is size-independent. In the mtc-QSAR-ANN model, *DGTI*08 and *DGTI*11 are the sixth and fourth most significant *D*[*GTI*]*cj* descriptors, respectively. In the case of *DGTI*10 (the eighth most important), its increase describes the shape of the molecules[64] by augmenting the number of linear fragments (propane-like skeletons). Last, we have *DGTI*15, which is the third most influential among all the *D*[*GTI*]*cj* descriptors in the mtc-QSAR-ANN model and a measure of the flexibility of the molecules.[65] In the context of the present work, the diminution of the value of *DGTI*15 is expected, particularly by increasing the number of fused rings.

Bearing in mind all the ideas mentioned when explaining the different *D*[*GTI*]*cj* descriptors, their joint interpretation can be summarized in the following manner. Electronegative atoms should be distributed through the entire molecular structure of a chemical, with an emphasis on functional groups based on nitrogen and oxygen (a sulfoxide group is allowed mainly in the periphery of a molecule as a replacement of an amide group). Ramifications should be avoided, and if present, they should preferably appear in one of the extremes of any molecule. The use of two fused ring systems separated by a linker containing four or five atoms (with at least two of them being aliphatic carbons) is the most important aspect. In this sense, each fused ring system should be formed by two heteroaromatic rings or a combination of a benzene ring with an aliphatic portion. Benzene rings, when present in the molecules, should present substitutions in multiple positions, with nitrogen, oxygen, and/or fluorine as the preferred atoms to substitute the hydrogen atoms.

2.3. Virtual Design of New Chemical with Pan-Antiviral and Anti-CS Profiles. By strictly following the joint interpretation of the 15 *D*[*GTI*]*cj* descriptors present in the mtc-QSAR-ANN model, we designed eight structurally related molecules (Figure 3). In doing so, we connected and/or fused some key molecular fragments,[66] i.e., those whose presence was suggested as desirable for the favorable variation of the values of most of the *D*[*GTI*]*cj* descriptors. Such

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

Table 5. Summary of the Predictions Performed by the mtc-QSAR-ANN Model for the Designed Molecules

cj[a,b]	DP-001	DP-002	DP-003	DP-004	DP-005	DP-006	DP-007	DP-008
c1	52.88	60.88	24.08	29.62	71.47	86.00	70.60	71.84
c2	88.40	91.62	71.30	78.93	96.18	99.06	95.29	93.36
c3	90.56	91.98	73.59	78.25	94.58	95.56	89.16	89.46
c4	75.59	78.16	98.34	98.56	88.62	86.73	96.56	96.56
c5	95.54	96.10	98.93	99.09	98.04	98.28	99.23	99.20
c6	65.55	71.68	61.22	68.75	63.53	82.51	79.92	74.79
c7	87.66	84.65	89.23	86.12	76.52	63.78	51.61	57.04
c8	12.44	17.24	9.92	14.98	33.60	76.73	83.15	78.64
c9	80.37	80.37	88.57	90.72	90.01	95.24	98.51	98.41
c10	70.77	77.26	43.63	50.75	84.86	92.18	82.03	83.10
c11	88.97	92.17	73.60	80.72	96.22	98.95	93.86	91.29
c12	62.41	67.01	97.46	97.89	82.87	81.14	94.97	94.70
c13	42.66	50.53	55.69	64.14	46.43	71.93	75.12	68.21
c14	18.20	25.63	17.83	26.53	49.93	86.47	90.15	86.64
c15	80.27	81.67	91.00	93.23	92.40	96.54	98.81	98.63
c16	9.93	5.63	41.45	24.99	8.38	11.35	60.80	73.98
c17	98.31	98.08	96.43	96.55	99.04	99.03	97.20	97.20
c18	88.70	89.80	89.77	91.90	94.64	95.43	96.21	95.50
c19	59.44	66.91	18.14	22.38	60.12	74.77	28.50	23.29
c20	44.64	42.04	46.34	44.05	48.14	54.20	55.20	54.41
c21	0.43	0.31	7.54	4.32	0.07	0.10	4.16	7.26
c22	54.11	55.17	61.31	61.79	62.25	73.36	71.89	70.13
c23	92.53	90.32	82.20	77.51	86.73	75.53	50.58	58.71

[a]This refers to the different experimental conditions as reported in Table 1. [b]The numbers in this table are the predicted values of probability for each molecule to be considered active. In the Supporting Information (Table S10), these probability values appear in a column named Prob. (%)Act.

Table 6. Physicochemical Properties Suggesting the Druglikeness of the Designed Molecules

ID[a]	HD	HA	MW	*M* log *P*	*A* log *P*	AMR	NAT	NRB	TPSA
DP-001	3	9	448.51	1.247	1.614	112.72	52	6	133.17
DP-002	3	9	448.51	1.247	1.614	112.72	52	6	133.17
DP-003	4	10	429.44	0.914	1.548	106.65	51	6	125.99
DP-004	4	10	429.44	0.914	1.548	106.65	51	6	125.99
DP-005	4	10	450.51	1.137	1.330	112.99	51	5	127.77
DP-006	3	10	451.49	1.137	1.977	111.04	50	6	124.97
DP-007	4	10	430.45	1.762	1.904	108.41	51	6	108.56
DP-008	4	10	430.45	1.762	1.904	108.41	51	6	108.56

[a]In the table, the abbreviations have the following meanings: the number of atoms behaving as hydrogen bond donors (HD), the number of atoms acting as hydrogen bond acceptors (HA), the molecular weight (MW), the logarithm of Moriguchi's octanol/water partition coefficient (*M* log *P*), the logarithm of Ghose−Crippen's octanol/water partition coefficient (*A* log *P*), Ghose−Crippen's molar refractivity (AMR), the number of atoms (NAT), the number of rotatable bonds (NRB), and the topological polar surface area (PSA).

fragments included fused ring systems, as well as amide, sulfoxide, carbamate, and urea groups. We also decorated the designed molecules with certain atoms/functional groups such as fluor and hydroxymethyl (both in the periphery of the molecule), as well as nitrogen atoms from secondary amines.

The designed molecules were predicted by our mtc-QSAR-ANN model to confirm that the virtual design was correctly performed. Such predictions are depicted in Table 5. The experimental conditions from *c1* to *c15* involve the predicted antiviral activities against the different viral strains while those in the intervals *c16*−*c18* and *c19*−*c23* are based on the inhibitory potencies against caspase-1 and TNF-alpha, respectively. All the data regarding the eight designed molecules can be found in detail in the Supporting Information (Tables S8−S11).

The numbers in Table 5 reflect the probabilities (predicted by the mtc-QSAR-ANN model) of the designed molecules to be considered active. Because most of the probability values are higher than 50%, we can infer that the designed molecules seem to have pan-antiviral activity while also inhibiting the CS-related proteins caspase-1 and TNF-alpha. From a physicochemical and structural point of view, we can say that fragments such as the two systems of fused rings (each of them with one polar part and one hydrophobic region), the presence and location of the two fluorine atoms, and even the hydroxymethyl group (from DP-001 to DP-004) positively account for the dual pan-antiviral and anti-CS profiles of all the designed molecules. In any case, notice that DP-006, DP-007, and DP-008 are the most desirable molecules since they were predicted as active in a larger number of experimental conditions *cj* and with the highest probabilities in most of these conditions. This indicates that the nitrogen atom from a secondary amine is necessary as well as the seemingly more correct placement of the amide and sulfoxide groups. Also, the carbamate group is slightly preferred over the urea moiety. We would like to emphasize that because the eight designed

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

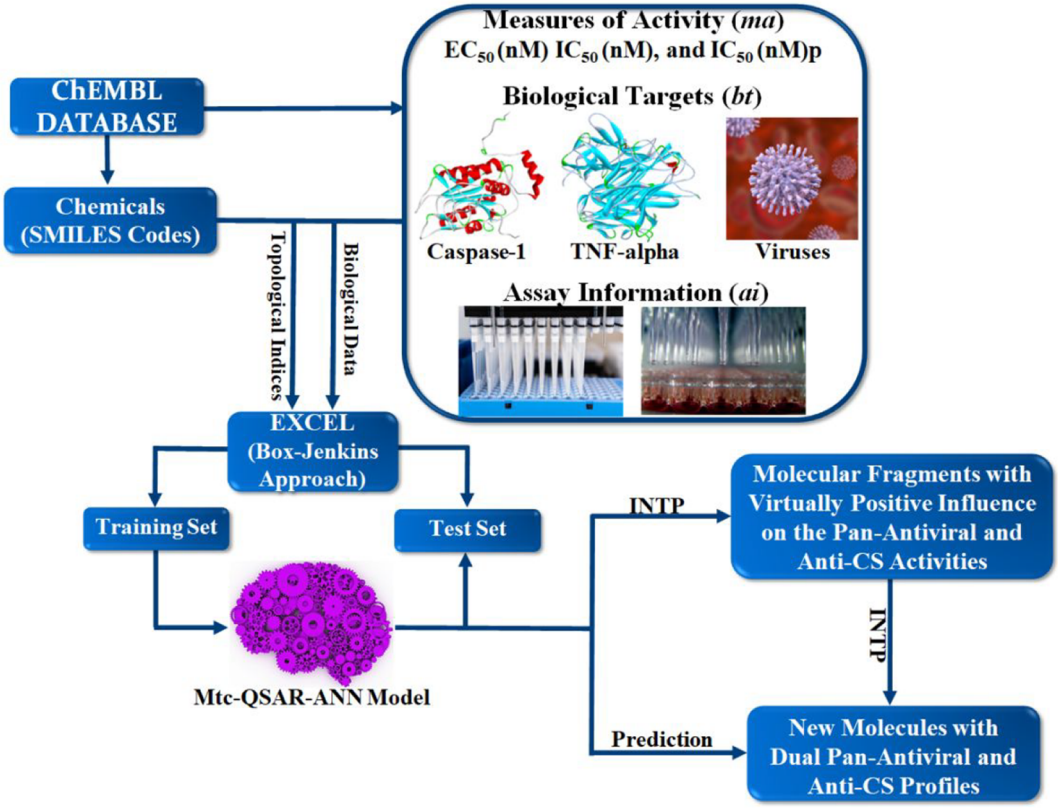


Figure 4. Development and use of an mtc-QSAR-ANN model. The $D[GTI]cj$ descriptors were calculated by applying the Box-Jenkins approach; such calculations were carried out in Microsoft Excel. Before finding the mtc-QSAR-ANN model, the data set was randomly split into training and test sets, which accounted for 75% and 25% of the data set, respectively. The abbreviation "INTP" signifies the physicochemical and structural interpretations of the $D[GTI]cj$ descriptors.

molecules were predicted by the mtc-QSAR-ANN model against the 23 experimental conditions for *cj*, 184 predictions in total were performed; in all the predictions, the designed molecules fell within the applicability domain of the mtc-QSAR-ANN model.

We also examined the druglikeness of the designed molecules by calculating a series of physicochemical properties[67] that are depicted in Table 6. The purpose here was to determine whether these molecules simultaneously complied with Lipinski's rule of five,[68] Ghose's filter,[69] and Veber's recommendations.[70] The values of the physicochemical properties of the designed molecules were in agreement with the corresponding cutoff values of the physicochemical properties established by the three aforementioned approaches, confirming their adequate druglikeness.

Last, intending to assess the novelty of the designed molecules, we performed a search in prestigious external databases such as ChEMBL[71] and ZINC.[72] The idea was to investigate in these databases the existence of chemicals that could structurally resemble our designed molecules. By applying a similarity cutoff value of 85%, we did not find any molecule similar to ours.

3. CONCLUSION

Computational methods have the potential to accelerate the discovery of therapeutic chemicals in the context of antiviral research. However, given the complexity of the infections caused by respiratory viruses, such methods should go beyond the single task of performing virtual screening by considering only one viral target. The mtc-QSAR-ANN model developed by us is a confirmation that chemical and biological data can be successfully integrated, allowing the simultaneous prediction of pan-antiviral activity against different respiratory viruses and inhibitory potency against CS-related proteins. We have also demonstrated the importance of the physicochemical and structural interpretations of the molecular descriptors, which enabled the use of the mtc-QSAR-ANN model as a tool to design novel molecules with virtually dual pan-antiviral and anti-CS profiles. The methodology underpinning the creation and application of our mtc-QSAR-ANN model opens encouraging opportunities for the *in silico* design of chemicals with the desired properties.

4. MATERIALS AND METHODS

The development and application of our mtc-QSAR-ANN model are illustrated in Figure 4. All the chemical and biological data used in this work were retrieved from version 29 of the ChEMBL database.[71,73] The curation of our data set was carried out according to certain previously reported guidelines.[35,40,66,74]

By using the software named MODESLAB v1.5,[75] we calculated (using the SMILES codes stored in a txt file) the following topology-based molecular descriptors (*TI*): atom connectivity indices, bond-based connectivity indices, spectral moments of the bond adjacency matrix, Kier's shape descriptors, and Kier's flexibility index and other classical topological indices. Other steps such as the calculation of a set of size-independent descriptors (*NTI*), the application of the

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130

Box–Jenkins approach to obtain the multicondition descriptors $D[GTI]cj$, the split of the data set of the training and test sets, the selection of the most adequate $D[GTI]cj$ descriptors, and the creation of the mtc-QSAR-ANN model were performed by strictly following a recent work.[76] Therefore, we will mention here only specific aspects.

When calculating the $D[GTI]cj$ descriptors,[76] the *a priori* probabilities $ps(cj)$ were defined according to the following mathematical formalism:

$$ps(cj) = \frac{n_A(cj)}{N_T(cj)} \quad (1)$$

In eq 1, $n_A(cj)$ and $N_T(cj)$ are the numbers of chemicals/cases annotated as active and the total number of cases, respectively. Both numbers (calculated by considering only chemicals/cases in the training set) depend on a specific element of the experimental condition cj. For instance, if $cj = ma$, then, $ps(ma)$, $n_A(ma)$, and $N_T(ma)$ correspond to the *a priori* probability, the number of active chemicals/cases, and the total number of chemicals/cases, all of them by considering only the element named measure of activity (ma). The same equation was applied separately to the elements named biological target (bt) and assay information (ai).

When selecting the most appropriate $D[GTI]cj$ descriptors using the software IMMAN v1.0,[77] we computed Jeffreys's information index (100 bins) to rank them according to their potential discriminatory power. Then, we performed a correlation analysis, keeping only those $D[GTI]cj$ descriptors with pairwise correlation values in the interval $-0.7 < PCC < +0.7$ (PCC was for the Pearson's correlation coefficient). To find the best mtc-QSAR-ANN model, we employed the ANN package of the software STATISTICA v13.5.0.17.[78] In doing so, we utilized a customized configuration [strategy for creating predictive models: automated network search; network types, MLP; minimum hidden units (neurons), 15; maximum hidden units (neurons), 70; networks to train, 3000; networks to retain, 50; activation functions for the hidden layer, logistic and hyperbolic tangent; activation function for the output layer, logistic, hyperbolic tangent, and softmax (the latter was implicitly used by default)]. The options regarding weight decay and initialization were left inactivated. We considered only those MLP networks with the number of epochs equal to or less than 400. In the end, the best MLP network (most suitable mtc-QSAR-ANN model) was the one exhibiting the highest values of the local measures $[Sn\ (\%)]ma$, $[Sp\ (\%)]ma$, $[Sn\ (\%)]bt$, $[Sp\ (\%)]bt$, $[Sn\ (\%)]ai$, and $[Sp\ (\%)]ai$. We would like to emphasize that our choice of using ANN via MLP as the machine learning algorithm with the aforementioned configuration is based on our previous experience in the creation and application of PTML models with good statistical quality and predictive power.[34,66,76,79]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.2c03363.

Cover page, title of the manuscript as well as names, affiliations, and contacts of the authors; Tables S1 and S2, Chemical and biological data, topological indices, averages, and standard deviation values; Tables S3–S7, $D[GTI]cj$ descriptors, classification results, local measures of performance, and applicability domain of the mtc-QSAR-ANN model; and Tables S8–S11, topological indices, $D[GTI]cj$ descriptors, prediction results, and assessment of the applicability for the designed molecules (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Alejandro Speck-Planche − *Grupo de Química Computacional y Teórica (QCT-USFQ), Departamento de Ingeniería Química, Universidad San Francisco de Quito, Quito 170901, Ecuador;* orcid.org/0000-0002-9544-9016; Email: alejspivanovich@yahoo.es, alejspivanovich@gmail.com

Author

Valeria V. Kleandrova − *Laboratory of Fundamental and Applied Research of Quality and Technology of Food Production, Moscow State University of Food Production, 125080 Moscow, Russian Federation;* orcid.org/0000-0002-1928-853X

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c03363

Author Contributions

Conceptualization, A.S.-P.; methodology, A.S.-P.; software, A.S.-P. and V.V.K.; 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. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Notes

The authors declare no competing financial interest.

■ REFERENCES

(1) Dong, E.; Du, H.; Gardner, L. An Interactive Web-Based Dashboard to Track Covid-19 in Real Time. *Lancet Infect. Dis.* **2020**, *20*, 533–534.

(2) Zhong, N. S.; Zheng, B. J.; Li, Y. M.; Poon; Xie, Z. H.; Chan, K. H.; Li, P. H.; Tan, S. Y.; Chang, Q.; Xie, J. P.; Liu, X. Q.; Xu, J.; Li, D. X.; Yuen, K. Y.; Peiris; Guan, Y. Epidemiology and Cause of Severe Acute Respiratory Syndrome (Sars) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **2003**, *362*, 1353–1358.

(3) Chan-Yeung, M.; Xu, R. H. Sars: Epidemiology. *Respirology* **2003**, *8*, S9–S14.

(4) Dunning, J.; Baillie, J. K.; Cao, B.; Hayden, F. G. International Severe Acute, R.; Emerging Infection, C. Antiviral Combinations for Severe Influenza. *Lancet Infect. Dis.* **2014**, *14*, 1259–1270.

(5) Sharma, L.; Rebaza, A.; Dela Cruz, C. S. When "B" Becomes "A": The Emerging Threat of Influenza B Virus. *Eur. Respir. J.* **2019**, *54*, 1901325.

(6) Zheng, Z.; Pitzer, V. E.; Shapiro, E. D.; Bont, L. J.; Weinberger, D. M. Estimation of the Timing and Intensity of Reemergence of Respiratory Syncytial Virus Following the Covid-19 Pandemic in the Us. *JAMA Netw. Open* **2021**, *4*, No. e2141779.

(7) Lai, C. C.; Wang, C. Y.; Hsueh, P. R. Co-Infections among Patients with Covid-19: The Need for Combination Therapy with Non-Anti-Sars-Cov-2 Agents? *J. Microbiol. Immunol. Infect.* **2020**, *53*, 505–512.

(8) Swets, M. C.; Russell, C. D.; Harrison, E. M.; Docherty, A. B.; Lone, N.; Girvan, M.; Hardwick, H. E.; ISARIC4C Investigators; Visser, L. G.; Openshaw, P. J. M.; Groeneveld, G. H.; Semple, M. G.;

Baillie, J. K. Sars-Cov-2 Co-Infection with Influenza Viruses, Respiratory Syncytial Virus, or Adenoviruses. *Lancet* **2022**, *399*, 1463−1464.
(9) Yoshida, L. M.; Suzuki, M.; Nguyen, H. A.; Le, M. N.; Dinh Vu, T.; Yoshino, H.; Schmidt, W. P.; Nguyen, T. T.; Le, H. T.; Morimoto, K.; Moriuchi, H.; Dang, D. A.; Ariyoshi, K. Respiratory Syncytial Virus: Co-Infection and Paediatric Lower Respiratory Tract Infections. *Eur. Respir. J.* **2013**, *42*, 461−469.
(10) Olbei, M.; Hautefort, I.; Modos, D.; Treveil, A.; Poletti, M.; Gul, L.; Shannon-Lowe, C. D.; Korcsmaros, T. Sars-Cov-2 Causes a Different Cytokine Response Compared to Other Cytokine Storm-Causing Respiratory Viruses in Severely Ill Patients. *Front. Immunol.* **2021**, *12*, 629193.
(11) Gu, Y.; Zuo, X.; Zhang, S.; Ouyang, Z.; Jiang, S.; Wang, F.; Wang, G. The Mechanism Behind Influenza Virus Cytokine Storm. *Viruses* **2021**, *13*, 1362.
(12) Bohmwald, K.; Galvez, N. M. S.; Canedo-Marroquin, G.; Pizarro-Ortega, M. S.; Andrade-Parra, C.; Gomez-Santander, F.; Kalergis, A. M. Contribution of Cytokines to Tissue Damage During Human Respiratory Syncytial Virus Infection. *Front. Immunol.* **2019**, *10*, 452.
(13) Dinarello, C. A. Interleukin 1 and Interleukin 18 as Mediators of Inflammation and the Aging Process. *Am. J. Clin. Nutr.* **2006**, *83*, 447S−455S.
(14) 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.
(15) 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. *Joint Bone Spine* **2016**, *83*, 282−289.
(16) Jang, D. I.; Lee, A. H.; Shin, H. Y.; Song, H. R.; Park, J. H.; Kang, T. B.; Lee, S. R.; Yang, S. H. The Role of Tumor Necrosis Factor Alpha (Tnf-Alpha) in Autoimmune Disease and Current Tnf-Alpha Inhibitors in Therapeutics. *Int. J. Mol. Sci.* **2021**, *22*, 2719.
(17) 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.
(18) Tanabe, K.; Matsushima-Nishiwaki, R.; Yamaguchi, S.; Iida, H.; Dohi, S.; Kozawa, O. Mechanisms of Tumor Necrosis Factor-A-Induced Interleukin-6 Synthesis in Glioma Cells. *J. Neuroinflammation* **2010**, *7*, 16.
(19) 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.
(20) Kleandrova, V. V.; Scotti, M. T.; Speck-Planche, A. Indirect-Acting Pan-Antivirals Vs. Respiratory Viruses: A Fresh Perspective on Computational Multi-Target Drug Discovery. *Curr. Top. Med. Chem.* **2021**, *21*, 2687−2693.
(21) Schaduangrat, N.; Lampa, S.; Simeon, S.; Gleeson, M. P.; Spjuth, O.; Nantasenamat, C. Towards Reproducible Computational Drug Discovery. *J. Cheminformatics* **2020**, *12*, 9.
(22) Toropov, A. A.; Toropova, A. P.; Veselinovic, A. M.; Leszczynska, D.; Leszczynski, J. Sars-Cov M(Pro) Inhibitory Activity of Aromatic Disulfide Compounds: Qsar Model. *J. Biomol. Struct. Dyn.* **2022**, *40*, 780−786.
(23) Jawarkar, R. D.; Bakal, R. L.; Zaki, M. E. A.; Al-Hussain, S.; Ghosh, A.; Gandhi, A.; Mukerjee, N.; Samad, A.; Masand, V. H.; Lewaa, I. Qsar Based Virtual Screening Derived Identification of a Novel Hit as a Sars Cov-229e 3cl(Pro) Inhibitor: Ga-Mlr Qsar Modeling Supported by Molecular Docking, Molecular Dynamics Simulation and Mmgbsa Calculation Approaches. *Arab. J. Chem.* **2022**, *15*, 103499.
(24) Cox, R. M.; Toots, M.; Yoon, J. J.; Sourimant, J.; Ludeke, B.; Fearns, R.; Bourque, E.; Patti, J.; Lee, E.; Vernachio, J.; Plemper, R. K. Development of an Allosteric Inhibitor Class Blocking Rna Elongation by the Respiratory Syncytial Virus Polymerase Complex. *J. Biol. Chem.* **2018**, *293*, 16761−16777.
(25) Rohini, K.; Shanthi, V. Hyphenated 3d-Qsar Statistical Model-Drug Repurposing Analysis for the Identification of Potent Neuraminidase Inhibitor. *Cell Biochem. Biophys.* **2018**, *76*, 357−376.
(26) Rohini, K.; Ramanathan, K.; Shanthi, V. Multi-Dimensional Screening Strategy for Drug Repurposing with Statistical Framework-a New Road to Influenza Drug Discovery. *Cell Biochem. Biophys.* **2019**, *77*, 319−333.
(27) Ghosh, A. K.; Takayama, J.; Aubin, Y.; Ratia, K.; Chaudhuri, R.; Baez, Y.; Sleeman, K.; Coughlin, M.; Nichols, D. B.; Mulhearn, D. C.; Prabhakar, B. S.; Baker, S. C.; Johnson, M. E.; Mesecar, A. D. Structure-Based Design, Synthesis, and Biological Evaluation of a Series of Novel and Reversible Inhibitors for the Severe Acute Respiratory Syndrome-Coronavirus Papain-Like Protease. *J. Med. Chem.* **2009**, *52*, 5228−5240.
(28) Muthusamy, K.; Kirubakaran, P.; Krishnasamy, G.; Thanashankar, R. R. Computational Insights into the Inhibition of Influenza Viruses by Rupestonic Acid Derivatives: Pharmacophore Modeling, 3d-Qsar, Comfa and Comsia Studies. *Comb. Chem. High Throughput Screen.* **2015**, *18*, 63−74.
(29) Chen, Z.-D.; Zhao, L.; Chen, H.-Y.; Gong, J.-N.; Chen, X.; Chen, C. Y.-C. A Novel Artificial Intelligence Protocol to Investigate Potential Leads for Parkinson's Disease. *RSC Adv.* **2020**, *10*, 22939−22958.
(30) Kumi, R. O.; Soremekun, O. S.; Issahaku, A. R.; Agoni, C.; Olotu, F. A.; Soliman, M. E. S. Exploring the Ring Potential of 2,4-Diaminopyrimidine Derivatives Towards the Identification of Novel Caspase-1 Inhibitors in Alzheimer's Disease Therapy. *J. Mol. Model.* **2020**, *26*, 68.
(31) Jade, D. D.; Pandey, R.; Kumar, R.; Gupta, D. Ligand-Based Pharmacophore Modeling of Tnf-Alpha to Design Novel Inhibitors Using Virtual Screening and Molecular Dynamics. *J. Biomol. Struct. Dyn.* **2022**, *40*, 1702−1718.
(32) Halim, S. A.; Sikandari, A. G.; Khan, A.; Wadood, A.; Fatmi, M. Q.; Csuk, R.; Al-Harrasi, A. Structure-Based Virtual Screening of Tumor Necrosis Factor-Alpha Inhibitors by Cheminformatics Approaches and Bio-Molecular Simulation. *Biomolecules* **2021**, *11*, 329.
(33) Gonzalez-Diaz, H.; Arrasate, S.; Gomez-SanJuan, A.; Sotomayor, N.; Lete, E.; Besada-Porto, L.; Ruso, J. M. General Theory for Multiple Input-Output Perturbations in Complex Molecular Systems. 1. Linear Qspr Electronegativity Models in Physical, Organic, and Medicinal Chemistry. *Curr. Top. Med. Chem.* **2013**, *13*, 1713−1741.
(34) Kleandrova, V. V.; Scotti, M. T.; Speck-Planche, A. Computational Drug Repurposing for Antituberculosis Therapy: Discovery of Multi-Strain Inhibitors. *Antibiotics (Basel)* **2021**, *10*, 1005.
(35) 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.
(36) Urista, D. V.; Carrue, D. B.; Otero, I.; Arrasate, S.; Quevedo-Tumailli, V. F.; Gestal, M.; Gonzalez-Diaz, H.; Munteanu, C. R. Prediction of Antimalarial Drug-Decorated Nanoparticle Delivery Systems with Random Forest Models. *Biology (Basel)* **2020**, *9*, 198.
(37) 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. Pharmaceutics* **2019**, *16*, 4200−4212.
(38) Nocedo-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.
(39) Kleandrova, V. V.; Scotti, M. T.; Scotti, L.; Nayarisseri, 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.
(40) 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.
(41) 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. Pharmaceutics* **2020**, *17*, 2612−2627.
(42) 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.
(43) 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.
(44) 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.
(45) 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.
(46) Kleandrova, V. V.; Ruso, J. M.; Speck-Planche, A.; Dias Soeiro Cordeiro, M. N. Enabling the Discovery and Virtual Screening of Potent and Safe Antimicrobial Peptides. Simultaneous Prediction of Antibacterial Activity and Cytotoxicity. *ACS Comb. Sci.* **2016**, *18*, 490−498.
(47) Prado-Prado, F.; Garcia-Mera, X.; Abeijon, P.; Alonso, N.; Caamano, O.; Yanez, M.; Garate, T.; Mezo, M.; Gonzalez-Warleta, M.; Muino, L.; Ubeira, F. M.; Gonzalez-Diaz, H. Using Entropy of Drug and Protein Graphs to Predict Fda Drug-Target Network: Theoretic-Experimental Study of Mao Inhibitors and Hemoglobin Peptides from Fasciola Hepatica. *Eur. J. Med. Chem.* **2011**, *46*, 1074−1094.
(48) Matthews, B. W. Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. *Biochim. Biophys. Acta* **1975**, *405*, 442−451.
(49) Sahigara, F.; Mansouri, K.; Ballabio, D.; Mauri, A.; Consonni, V.; Todeschini, R. Comparison of Different Approaches to Define the Applicability Domain of Qsar Models. *Molecules* **2012**, *17*, 4791−4810.
(50) 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*, 1661−1676.
(51) 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.
(52) Estrada, E. Physicochemical Interpretation of Molecular Connectivity Indices. *J. Phys. Chem. A* **2002**, *106*, 9085−9091.
(53) Randić, M.; Zupan, J. On Interpretation of Well-Known Topological Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 550−560.
(54) 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.
(55) 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.
(56) 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.
(57) 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.
(58) 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.
(59) Estrada, E.; Molina, E. Automatic Extraction of Structural Alerts for Predicting Chromosome Aberrations of Organic Compounds. *J. Mol. Graph. Model.* **2006**, *25*, 275−288.
(60) Estrada, E. Edge Adjacency Relationship and a Novel Topological Index Related to Molecular Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 31−33.
(61) 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.
(62) 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.
(63) Balaban, A. T. Distance Connectivity Index. *Chem. Phys. Lett.* **1982**, *89*, 399−404.
(64) Kier, L. B. Shape Indexes of Orders One and Three from Molecular Graphs. *Quant. Struct.-Act. Relat.* **1986**, *5*, 1−7.
(65) Kier, L. B. An Index of Flexibility from Molecular Shape Descriptors. *Prog. Clin. Biol. Res.* **1989**, *291*, 105−109.
(66) 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.
(67) Alvascience-Srl. *Alvadesc (Software for Molecular Descriptor Calculation)*, v1.0.22, 2020.
(68) 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 Delivery Rev.* **2001**, *46*, 3−26.
(69) 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.
(70) 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.
(71) Mendez, D.; Gaulton, A.; Bento, A. P.; Chambers, J.; De Veij, M.; Felix, E.; Magarinos, M. P.; Mosquera, J. F.; Mutowo, P.; Nowotka, M.; Gordillo-Maranon, M.; Hunter, F.; Junco, L.; Mugumbate, G.; Rodriguez-Lopez, M.; Atkinson, F.; Bosc, N.; Radoux, C. J.; Segura-Cabrera, A.; Hersey, A.; Leach, A. R. Chembl: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930−D940.
(72) 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.
(73) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. Chembl: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−1107.
(74) 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.
(75) Estrada, E.; Gutiérrez, Y. *Modeslab*, v1.5; Santiago de Compostela: Spain, 2002−2004.
(76) 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.

(77) 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.

(78) TIBCO-Software-Inc. *Statistica (Data Analysis Software System)*, v13.5.0.17; TIBCO-Software-Inc., Palo Alto, CA, 2018.

(79) Kleandrova, V. V.; Rojas-Vargas, J. A.; Scotti, M. T.; Speck-Planche, A. Ptml Modeling for Peptide Discovery: In Silico Design of Non-Hemolytic Peptides with Antihypertensive Activity. *Mol. Divers.* **2021**, DOI: 10.1007/s11030-021-10350-z.

https://doi.org/10.1021/acsomega.2c03363
ACS Omega 2022, 7, 32119−32130