

Article

DesMol2, an Effective Tool for the Construction of Molecular Libraries and Its Application to QSAR Using Molecular Topology

Inma García-Pereira ¹, Riccardo Zanni ², Maria Galvez-Llompart ^{2,3}, Jorge Galvez ² and Ramón García-Domenech ^{2,*}

¹ Institute of Robotics and Information and Communication Technologies (IRTIC), University of Valencia, 46100 Valencia, Spain; inmaculada.garcia-pereira@uv.es

² Molecular Topology and Drug Design Unit, Department of Physical Chemistry, University of Valencia, Avenida V.A. Estelles s/n, Burjassot, 46100 Valencia, Spain; riccardo.zanni@uv.es (R.Z.); maria.galvez@uv.es (M.G.-L.); Jorge.galvez@uv.es (J.G.)

³ Microbiology and Plant Pathology-Unit (CSIC Associated), Department of Microbiology, Faculty of Sciences, University of Malaga, 29071 Malaga, Spain

* Correspondence: ramon.garcia@uv.es; Tel.: +34-963-544-291

Academic Editor: Kok Hwa Lim

Received: 17 January 2019; Accepted: 15 February 2019; Published: 18 February 2019



Abstract: A web application, DesMol2, which offers two main functionalities, is presented: the construction of molecular libraries and the calculation of topological indices. These functionalities are explained through a practical example of research of active molecules to the formylpeptide receptor (FPR), a receptor associated with chronic inflammation in systemic amyloidosis and Alzheimer's disease. Starting from a data(base) of 106 dioxopiperazine pyrrolidin piperazine derivatives and their respective constant values of binding affinity to FPR, multilinear regression and discriminant analyses are performed to calculate several predictive topological-mathematical models. Next, using the DesMol2 application, a molecular library consisting of 6,120 molecules is built and performed for each predictive model. The best potential active candidates are selected and compared with results from other previous works.

Keywords: DesMol2; topology descriptors; molecular libraries; multilinear regression analysis; linear discriminant analysis; Alzheimer's disease

1. Introduction

Nowadays, one of the biggest challenges in chemistry is to associate a given property (toxicity, biodegradation, pharmacological activity) to a specific compound. For instance, if given a certain chemical structure, being able to determine if it will show a certain experimental property. Researchers, who search for molecules with certain physicochemical, biological or pharmacological properties, have large molecular databases at their disposal. In fact, there are currently more than 144 million chemical compounds cataloged with their CAS registration number [1]. In addition to different commercial molecular databases currently available on the internet (ChemIDplus, DrugBank, Zinc, ChemBank, Specs, Molbase, etc.), along with the information that can be obtained from bibliographic databases (SciFinder, PubMed, Scopus, etc.), researchers can build new molecules thanks to combinatorial chemistry and virtual synthesis. Current software allows the creation of a new molecule in a matter of seconds, without the need of a laboratory to perform the real synthesis. All this requires is the use of quick and effective filtering methodologies and the selection of the most interesting compounds in each field of research and development. QSAR (*Quantitative Structure-Activity Relationship*) modeling

techniques represents one of the main computational tools used address this challenge and the process of designing and searching for new compounds.

The core of any QSAR model lies in molecular descriptors. These descriptors are made up of a series of numerical values that identify the molecule along with a specific physical, chemical or biological property. Depending on the type of information that is parameterized, one can differentiate between constitutional, topological (including topochemical) electrostatic, geometric and quantum indices [2].

In particular, the structural characterization of a molecule through its topological indices is an extremely useful tool in QSAR analysis. Molecular topology which converts a molecule into a graph whose vertices are the atoms and whose edges are the bonds, is used to calculate these indices [3]. A proof of molecular topology are the results obtained in the design of new chemical compounds in areas as diverse as cancer [4], malaria [5], Alzheimer [6], mosquito repellents [7], green chemistry [8,9], etc.

Currently, several free access computer tools allow the construction of molecular libraries. The most complete tools are usually desktop applications. Among them are SmiLib v2.0 [10], ChemT [11] and Library Synthesizer [12]. The main weakness when using SmiLib or ChemT is that they do not verify if the SMILES (Simplified Molecular Input Line Entry Specification) codes entered by the user are correct, so that the resulting libraries may contain wrongly identified molecules. On the other hand, Library Synthesizer does not allow the introduction of molecules by means of SMILES code, forcing the user to draw them instead. This can be a very tedious process when the lists of substituents are large. In addition, all three programs have restrictions when introducing the different molecular fragments because they limit the quantity or the combinations that can be made between them. There are also other web applications whose purpose is the construction of molecule libraries. However, their functionalities are limited. For example, cheminfo.org offers a virtual combinatorial library [13], and e-LEA3D has its own combinatorial library design tool [14]. In both cases, the user is forced to draw the different molecular fragments since they do not allow direct entry of SMILES code.

As a result, DesMol2 provides significant advantages over the other leading software. For example, it allows the introduction of more than one base and lists of substituents for each anchor point at a time, and it does not matter where the anchor point of the substituent is placed. It also verifies that the introduced SMILES codes correspond to valid molecular fragments and re-numbers the cycles of the generated molecules to ensure that they are correct.

In this paper, a web application, DesMol2 [15] is used to combine two functionalities in a single environment. Firstly, it constructs molecular libraries and calculates topological indices. These operations are carried out without the need for any other software, supporting different formats of chemical files, offering the output in compatible formats with classic QSAR analysis programs and automating the creation of new molecules.

One of the biggest advantages of DesMol2 compared to other software with similar functionalities is that it internally checks each SMILES code entered by the user to verify that it is correct. If so, it accepts the fragment to be able to process it later. If it detects an anomaly, it informs the user so any corrections can be made. Similarly, DesMol2 only builds correct molecules, unlike other programs that simply bind SMILES code fragments without verifying whether the result is a well-constructed molecule or not. Our software is using The Chemistry Development Kit (CDK) to carry out this process; this is an open source library with LGPL license.

The steps to follow are:

Using DesMol2, design a complete library of molecules from the SMILES codes of the bases and the substituents defined by Besalú [16].

Calculate with DesMol2 the topological indices of each of the generated molecules.

Build the topological model of activity prediction against the FPR receptors.

Apply the topological model to the library and select the potentially most active molecules.

Compare the results with those reported in other works.

2. Results and Discussion

With the help of Statistica software, Lineal Discriminant Analysis (LDA) was performed with the training group, using the grouping variable (see Table S3, Clas(obs)) as the dependent variable and the topological indices calculated by DesMol2 as independent variables.

The discriminant function (DF) selected was:

$$DF = 245.5 - 296.2C^1\chi + 50.2D^4\chi_{pc} - 0.0034W + 4.4V4 \quad (1)$$

$$N = 85 \quad \lambda = 0.421 \quad F(4,80) = 27.5 \quad p < 0.00001$$

The DF is statistically significant, both for the equation as a whole, $p < 0.00001$, and for each of the variables that enter the model, $p < 0.004$.

The topological descriptors $C^1\chi$, $D^4\chi_{pc}$ and $V4$ provide fundamental topological information related to molecular branching. The model also includes the Wiener index.

When the DF function is applied to each compound, it is classified as active (A) if $DF > 0$ or inactive (I) if $DF < 0$.

Table S3 shows the DF values obtained for each compound together with the activity probability (Prob (A)) and the predicted classification (Clas (pred)) for both the training group and the external validation group. The molecules marked with an asterisk represent classification errors.

Figure 1 shows the sensitivity (percentage of actives correctly classified), specificity (percentage of inactives correctly classified) and accuracy (percentage global correctly classified) for each data set. In both sets, specificity is higher than 86.5%, which reduces to the maximum the possibility of false assets, while the sensitivity is higher than 80.5%.

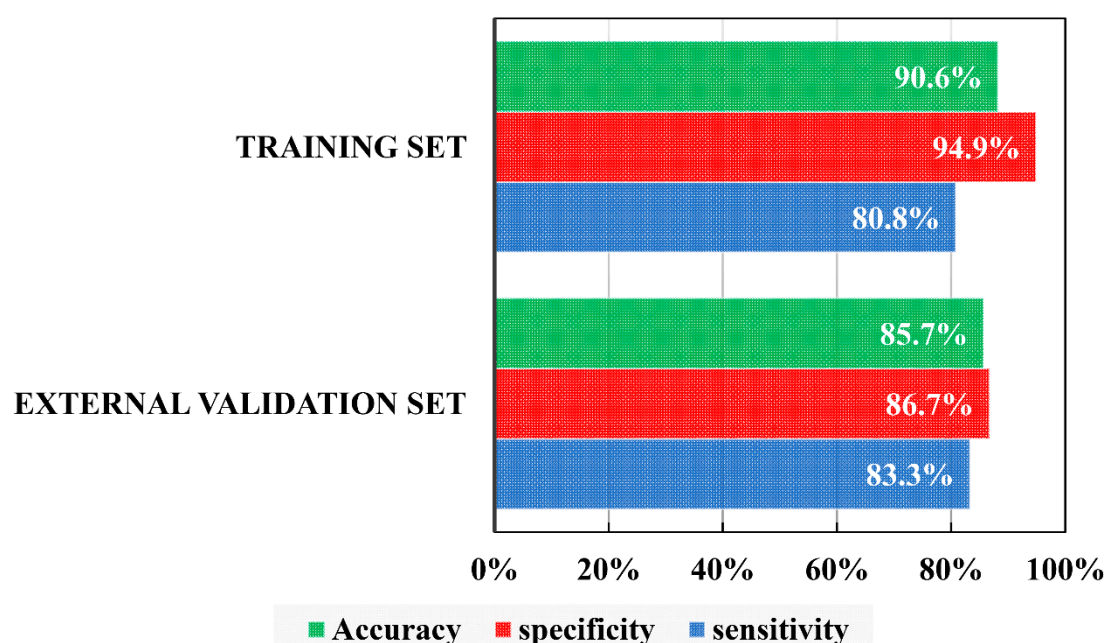


Figure 1. Percentages of correct classification obtained from the LDA (Linear Discriminant Analysis) for each set.

To determine the domain of applicability of the selected discriminant function, DF, the corresponding activity distribution diagram was created, which is shown in Figure 2.

This diagram represents the expectation (probability) of finding active or inactive compounds for different DF values [17]. The red bars represent the active group ($\text{Log}k_i < 2.60$) and the blue bars represent the inactive group ($\text{Log}k_i > 2.60$). Thus, the useful DF range for classifying the compounds

as active is between 0 and 8 such that a compound will be classified as active if DF falls within that range. Values of $DF < 0$ classify the compound as inactive, while values of $DF > 8$ or $DF < -8$ will be outside the domain of applicability of DF and will, therefore, be unclassified.

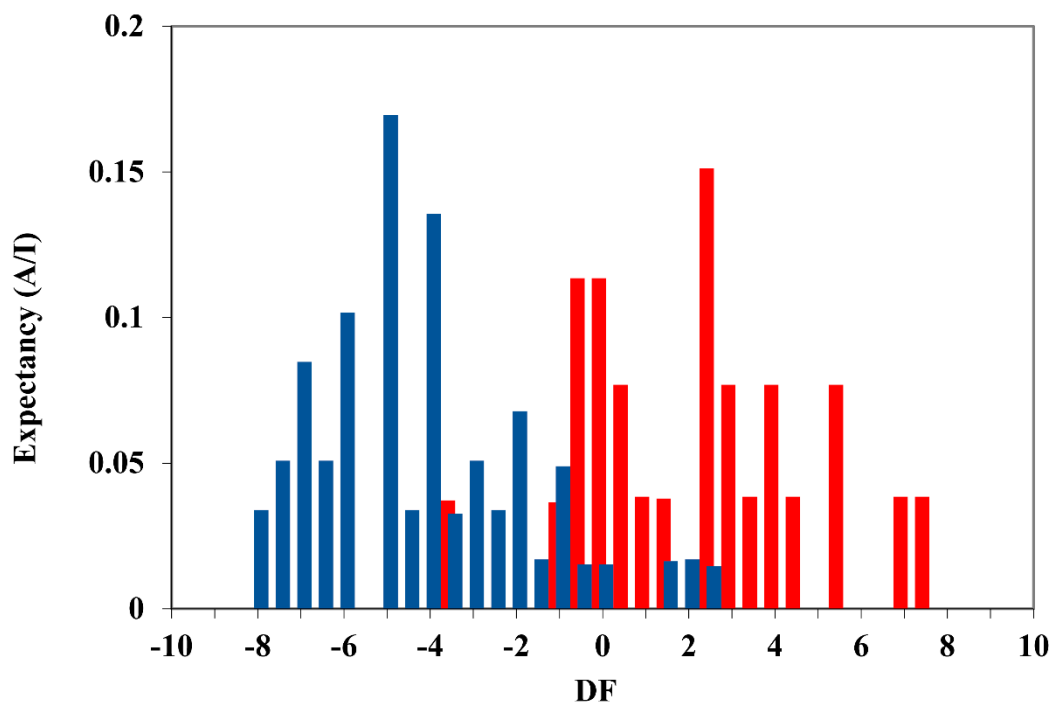


Figure 2. Pharmacological distribution diagram for bonding affinity, k_i , (red columns represents activity, and blue columns inactivity).

Once the mathematical model of classification (DF) has been selected, as well as the domain of application, the next step is to apply it to the rest of the previously constructed library. As previously mentioned, the topological indices have been calculated for each molecule with DesMol2. Next, the equation DF is applied to them. All molecules with DF values between 0 and 8 are classified as potentially active. The result was 1403 of a total of 6120, which represents approximately 23% of the bookstore. These 1403 molecules have been classified as active, with a theoretical value of $\text{Log}k_i < 2.60$, which represents activity at concentrations lower than 400 nM (see Table S5). To predict the $\text{Log}k_i$ value of each of these molecules, it will be necessary to perform the corresponding multilinear regression analysis (MLRA). In this case, experimental $\text{Log}k_i$ values for 48 of the 106 compounds of the working group are available ($\text{Log}k_i$ values = 4 represent molecules with a concentration higher than 10,000 nM and have been eliminated from the MLRA). Taking the $\text{Log}k_i$ variable as the dependent and the topological indices as the independent variables, Statistica was used again to perform MLRA analysis. Since the indices of Equation (1) did not correctly predict the $\text{Log}k_i$ values, which was to be expected as the equation was discriminant in nature and not quantitative, new topological indices were calculated using the Dragon program [18].

The prediction equation selected was:

$$\text{Log}k_i = 76.50 + 0.000033ww - 9.50\text{MAXDN} - 15.68\text{BELm5} - 3.18\text{GGI10} - 5.12\text{VEA1} \quad (2)$$

$$N = 48 \quad R^2 = 0.875 \quad Q^2 = 0.841 \quad \text{SEE} = 0.376 \quad F = 58.9 \quad p < 0.00001$$

which is a multilinear equation with 5 variables: hyper detour index, ww , [19]; maximal electrotopological negative variation, MAXDN [20] (this descriptor can be related to the nucleophilicity of the molecule [21]); lowest eigenvalue $n^\circ 5$ of Burden matrix / weighted by atomic masses, BELm5 [22]; topological charge index of order 10, GGI10 [23], and eigenvector coefficient sum from

adjacency matrix, VEA1 [22]. All of them were classified as statistically significant ($p < 0.005$) with a high predictive power as indicated by the high value of the correlation coefficient, $R^2 = 0.875$, as well as the prediction coefficient obtained in the "leave-one-out" cross-validation, $Q^2 = 0.841$. The standard error of estimation, $SEE = 0.376$, corresponds to a value less than 10% of the interval in which the $\text{Log}k_i$ property is moving. Table S4 and Figure 3 show the prediction results obtained for each compound and for the set, respectively.

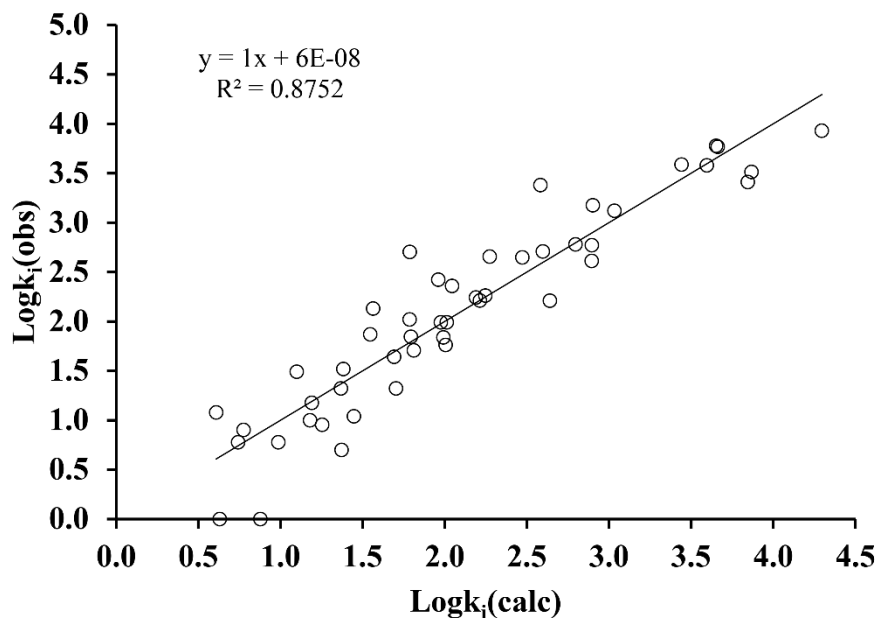


Figure 3. Graphical representation of $\text{Log}k_i$ (obs) versus $\text{Log}k_i$ (calc) obtained from Equation (2).

The randomness test performed at Equation (2) shows excellent stability thereof (see Figure 4).

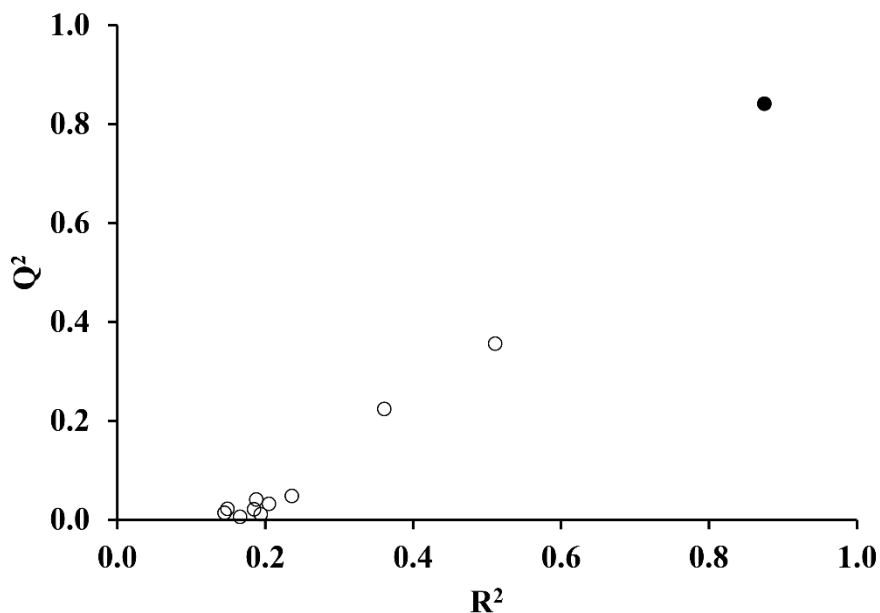


Figure 4. Randomness test performed on the selected regression model, Equation (2).

In view of the results obtained with the LDA and MLRA, a topological bonding prediction model affinity, k_i , can be established, consisting of the two functions: Equation (1) and Equation (2). This way, the library generated with DesMol2 (6120 molecules) can be used to select new potential active

molecules. The selection criteria were: A DF (Equation (1)) with values between 0 and 8 and $\text{Log}k_i$ (Equation (2)) with values lower than 2.60, which is equivalent to concentrations lower than 400 nM. Of the 1403 molecules theoretically classified by the DF as active, 785 had a predicted $\text{Log}k_i$ value < 2.60 , which is equivalent to approximately the 13% of the total library molecules (see Table S5, molecules marked in yellow, green and blue). If we limit the activity further and we impose a $\text{Log}k_i < 1.0$ ($k_i < 10$ nM), the number of highly active molecules is reduced to 87 (1.4% of the library) (see Table S5, molecules in yellow). Comparing these results with the ones by Besalú [16] using the SSIR (Superposing Significant Interaction Rules) method and the same library, it can be observed that of the 50 molecules selected by the SSIR method as the most active theoretical, 14 of them have been selected with our model with theoretical values of $\text{Log}k_i < 1.0$ (CAAA, CABG, CACE, CBAA, CBAF, CBBE, CBBG, CBCA, CBCG, CBCI, CBCL, CBCM, CBCP, CBHC) (see Table S5, molecules in yellow) and 28 with theoretical values of $\text{Log}k_i$ between 1.0 and 2.0 (CABB, CABD, CACA, CACF, CADC, CADH, CAFH, CAHC, CAHH, CAIC, CAIH, CBAD, CBAE, CBBB, CBCD, CBCJ, CBCK, CBCN, CBCO, CBCQ, CBDH, CBFC, CBFH, CBHH, CBIC, CBIH) (see Table S5, molecules in green). As can be seen, the most active compounds contain the isopropyl substituent in position R_1 ; 4-hydroxybenzyl or 2-naphthylmethyl in R_2 ; benzyl, 4-hydroxybenzyl or butyl in R_3 and different substituents in position R_4 . All this suggests that the substituents indicated above for positions R_1 and R_2 are crucial for the design of new active compounds against the FPR receptor. Figure 5 shows the chemical structure of some of the potentially most active molecules.

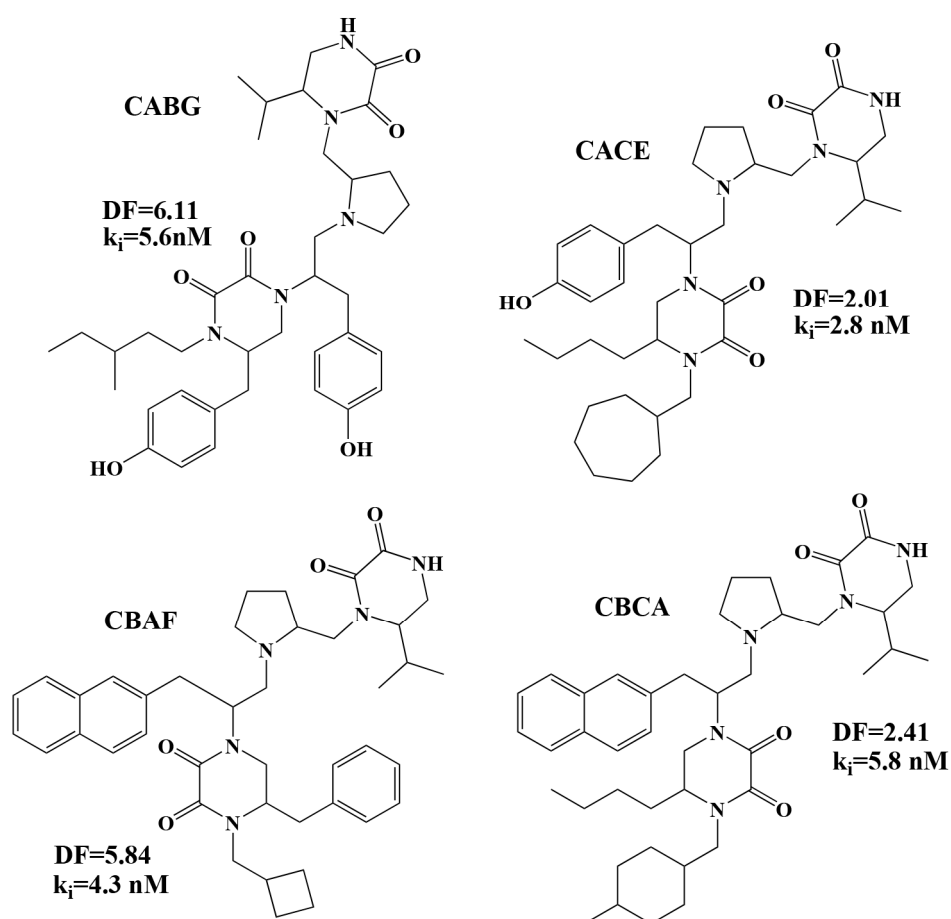


Figure 5. Chemical structure selected by the QSAR model with potential binding activity against the FPR receptor.

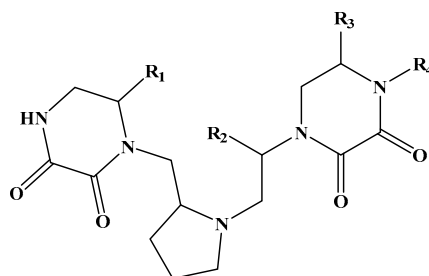
The next step in the search for new active compounds against RPF would be the synthesis of some of the molecules proposed in Figure 5 and Table S5 (molecules in yellow) and the realization of the corresponding pharmacological tests.

3. Materials and Methods

3.1. Data Set

The present work consists of a practical exercise using DesMol2 web application for the search of active molecules against the FPR receptors. It starts from the work by Besalú [16] that contains a library of 106 derivatives of dioxopiperazine pyrrolidin piperazine with their respective values of affinity constant of binding for the FPR, k_i , (see the Table S1 in the Supplementary Material). Besalú presents the SSIR (Superposing Significant Interaction Rules) method as a systematic procedure used to rank series of combinatorial analogues. Table 1 shows the substitution codifications along the four available molecular scaffold sites and Table S1, the codified and activity set of 106 compounds.

Table 1. Base structure and substituents used in the construction of the molecular library.



Code	R1	R2	R3	R4
A	2-naphthylmethyl [R]C1=CC2=CC=CC=C2C=C1	4-hydroxybenzyl OC1=CC=C(C=C1)[R]	benzyl [R]C1=CC=CC=C1	4-Methyl-1-cyclohexyl-methyl [R]C1CCC(C)CC1
B	propyl CC[R]	2-naphthylmethyl [R]C1=CC2=CC=CC=C2C=C1	4-hydroxybenzyl OC1=CC=C(C=C1)[R]	Cyclohexylpropyl [R]CCCC1CCCC1
C	isopropyl [R](C)(C)	cyclohexyl C1CCCCC1[R]	butyl [R]CCC	Cyclohexylmethyl [R]C1CCCCC1
D	butyl [R]CCC	propyl CC[R]	propyl CC[R]	Cyclopentylmethyl [R]C1CCCC1
E	benzyl [R]C1=CC=CC=C1	hydroxymethyl O[R]	S-phenyl [R]1=CC=CC=C1	Cycloheptylmethyl [R]C1CCCCC1
F		butyl [R]CCC	butyl [R]CCC	Cyclobutylmethyl [R]C1CCC1
G		benzyl [R]C1=CC=CC=C1	cyclohexyl C1CCCCC1[R]	3-Methylpentyl [R]CC(C)CC
H		isobutyl CC(C)[R]	benzyl [R]C1=CC=CC=C1	2-Biphenyl-4-yl-ethyl [R]CC(C=C1)=CC=C1C2=CC=CC=C2
I			propyl CC[R]	4-Tert-butyl-cyclohexylmethyl [R]C1CCC(C(C)(C)C)CC1
J				2-(3-Methoxyphenyl)-ethyl [R]CC1=CC(OC)=CC=C1
K				2-(4-Isobutylphenyl)-propyl [R]C(C1=CC=C(C(C)C)C=C1)C
L				m-Tolyethyl [R]CC1=CC(C)=CC=C1
M				p-Tolyethyl [R]CC1=CC=C(C)C=C1
N				2-(4-Methoxyphenyl)-ethyl [R]CC1=CC=C(OC)C=C1
O				2-(4-Ethoxyphenyl)-ethyl [R]CC1=CC=C(OCC)C=C1
P				Phenethyl [R]CC1=CC=CC=C1
Q				3-(3,4-Dimethoxyphenyl)-propyl [R]CCC1=CC=C(OC)C(OC)=C1

3.2. Building of the Virtual Molecular Library

The DesMol2 application meets the following functional requirements when building a molecular library:

- The application is capable of generating libraries of molecules by combining different bases, and substituents provided by the user.
- You can enter one or more bases and each of them can have one or more anchor points. For each anchor point, zero or more substituents can be entered. The following syntax indicates the anchor points: [R]. For each substituent, the atom that goes at the anchor point is also indicated.
- All possible combinations are made between bases and substituents.
- The user can graphically visualize each molecular structure, be it a base, a substituent or the result of the combination of both.
- The user is informed when an error occurs when entering the information incorrectly.
- The final result will be a *.smi file, so that each line of it represents a molecule through its code SMILES followed by a unique identifying name. The output file can be used for the calculation of molecular descriptors with other programs such as Dragon [18], PaDel [24], etc.

In the Supplementary Material section, Figure S2 shows a workflow diagram of DesMol2.

This work starts with a main structure and four anchor points: R₁, R₂, R₃ and R₄. For each of them, 5, 8, 9 and 17 substituents have been defined, respectively. Therefore, the library generated contains 6120 molecules (see Table S2). To build the library with DesMol2, the SMILES code was obtained from both the base and the substituents (see Table 1), and they were introduced in the DesMol2 application (see Figure S1). Once all the possible combinations between the base and substituents have been made, a *.smi file is obtained that contains the code SMILES of each generated molecule followed by the identification code (see Table S2). The *.smi file generated with the 6120 molecules library will be used to structurally identify each molecule through its corresponding topological indices, also calculated with DesMol2.

In this sense, DesMol2 is a very efficient application, since it is capable of building large libraries of molecules in a few seconds. For the data set chosen in this work, a library of 6120 molecules was built in just 6 s. For its part, the calculation of topological indices has also offered good results. To verify the computational efficiency of our web application, we have calculated the topological indexes of the 6120 molecules with DeMol2, Dragon and PaDel programs. To do this, a group of similar topological indices was selected in the three applications and the average time required for each molecule was obtained: 0.11 s for DesMol2, 0.05 for Dragon and 1.28 for PaDel. Keep in mind that DesMol2 is a web application while Dragon and PaDel are desktop applications. In addition, Dragon is a paid software, while PaDel and DesMol2 are free. Therefore, in view of the results obtained, we can affirm that our software has very competitive response times.

3.3. Molecular Descriptors

The DesMol2 application calculates the Kier and Hall topological connectivity indices up the fourth order, ${}^m\chi_t$ (indexes that evaluate fundamentally the topological assembly of molecules) [25], topological charge indices, Gi, Ji, (which evaluate the intramolecular charge transfer) [23], quotients and differences between the valence and non-valence connectivity indexes ($C^m\chi_\tau$, $D^m\chi_\tau$) [26], the Wiener index [27], and a group of constitutional indices (including the number of atoms, degree of branching, etc.) [26]. The recognized formats are: MDL Molfile (*.mol), Smiles (*.smi) and MDL DSfile (*.sdf). The generated topological indexes are formatted as an Excel file that contains two sheets: The first shows the calculated indices for each molecule, while the second contains the file information and unprocessed molecules.

3.4. Development of the QSAR Models

In this work, we have used linear discriminant analysis, LDA, and multilinear regression analysis, MLRA. The goal of the linear discriminant analysis (LDA) is to find a linear combination of variables that allow the discrimination between two or more categories or objects. In this case, the data set is formed by 106 molecules distributed in two groups: the active set, A, ($\text{Log } k_i < 2.60$, $k_i < 400$ nM) and the inactive set, I, ($\text{Log } k_i > 2.60$, $k_i > 400$ nM). LDA has been used to find the grouping descriptors that best discriminate the bonding affinity, k_i , with the Statistica 9.0 software [28]. The search for the discriminant function is carried out with the training group formed by 80% of the molecules (26 active and 59 inactive). The remaining 20%, selected at random, will serve as an external validation test (6 active and 15 inactive). In addition, as the experimental bonding affinity data of the active and inactive groups are quite unbalanced (26 active and 59 inactive), it is necessary to perform the LDA assigning different weights to each group depending on the existing imbalance. In this study, $p = 0.306$ and $p = 0.694$ were established for the active and inactive groups respectively. Multilinear regression analysis, MLRA, will be used to predict the activity of molecules classified as active by the LDA. $\text{Log } k_i$ values < 4.0 as the dependent variable and topological indices as independent variables were used respectively. In summary, information is available for 46 molecules.

The predictability, quality and robustness of the model selected can be verified by means of different types of criteria. In this study two strategies are adopted [29]:

- (a) Internal validation or cross-validation with leave-one-out (LOO) analysis. To do this, one compound of the set is extracted, and the model is recalculated using as a training set the remaining N-1 compounds. The property is then predicted for the removed element. This process is repeated for all the compounds of the set, obtaining a prediction for each of them. From the residual values obtained, the prediction coefficient, $R^2(\text{cv})$ or Q^2 is determined.
- (b) Data randomization or Y-scrambling. A randomization test can be analyzed to identify the possible existence of fortuitous correlations [30]. To do this, the values of the property of each compound are randomly permuted and linearly correlated with the topological descriptors.

4. Conclusions

DesMol2 software is presented as a tool available to researchers for the construction of molecular libraries and their subsequent implementation in QSAR analysis. Using SMILES code, large libraries can be designed consisting of one or more structural starting bases and different substituents or anchor points. The output format of the generated library is prepared for the calculation of topological descriptors, either with the DesMol2 itself or other existing software in the market and applied in the QSAR model search. DesMol2 is an open-access web application that, unlike other software with similar functionalities, provides great flexibility when introducing molecular fragments and always verifies the validity of both the introduced SMILES codes and the resulting molecules after making all possible combinations.

Supplementary Materials: All TIs used in the present study along with their values for every compound (training, test, and molecular library) are shown in the supporting information section.

Author Contributions: Writing-original draft, R.G.-D.; Software development, I.G.-P.; Computational data analysis; R.Z., M.G.-L. Writing-Review and editing, R.G.-D., R.Z., J.G.; Funding Acquisition, J.G., R.G.-D.

Funding: This research was funded by the MINECO (Spanish Ministry of Economy, Industry and Competitiveness) Project: "Desarrollo de nuevas herramientas para el control de oídios" (AGL2016-76216-C2-2-R).

Acknowledgments: The authors acknowledge the MINECO (Spanish Ministry of Economy, Industry and Competitiveness) Project: "Desarrollo de nuevas herramientas para el control de oídios" (AGL2016-76216-C2-2-R). I.G.P. acknowledges the Spanish Ministry of Science, Innovation and Universities (program: "University teacher formation") to carry out this study.

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

References

1. Chemical Abstracts Service. CAS Registry. The Gold Standard for Chemical Substance Information. Available online: <http://www.cas.org/content/chemical-substances> (accessed on 16 February 2019).
2. Basak, S.C. Role of mathematical chemodescriptors and proteomics based biodescriptors in drug discovery. *Drug Dev. Res.* **2011**, *72*, 225–233. [CrossRef]
3. Galvez, J.; Garcia-Domenech, R. On the contribution of molecular topology to drug design and discovery. *Curr. Comput.-Aided Drug Des.* **2010**, *6*, 252–268. [CrossRef] [PubMed]
4. Zanni, R.; Galvez-Llompert, M.; Morell, C.; Rodríguez-Henche, N.; Díaz-Laviada, I.; Recio-Iglesias, M.C.; Garcia-Domenech, R.; Galvez, J. Novel cancer chemotherapy hits by molecular topology: Dual akt and beta-catenin inhibitors. *PLoS ONE* **2015**, *10*, e0124244. [CrossRef] [PubMed]
5. Mahmoudi, N.; Garcia-Domenech, R.; Galvez, J.; Farhati, K.; Franetich, J.F.; Sauerwein, R.; Hannoun, L.; Derouin, F.; Danis, M.; Mazier, D. New active drugs against liver stages of Plasmodium predicted by molecular topology. *Antimicrob. Agents Chemother.* **2008**, *52*, 1215–1220. [CrossRef] [PubMed]
6. Wang, J.; Land, D.; Ono, K.; Galvez, J.; Zhao, W.; Vempati, P.; Steele, J.W.; Cheng, A.; Yamada, M.; Levine, S.; et al. Molecular topology as novel strategy for discovery of drugs with ab lowering and anti-aggregation dual activities for Alzheimer's disease. *PLoS ONE* **2014**, *9*, e92750.
7. Garcia-Domenech, R.; Aguilera, J.; El Moncef, A.; Pocovi, S.; Galvez, J. Application of molecular topology to the prediction of mosquito repellents of a group of terpenoid compounds. *Mol. Divers.* **2010**, *14*, 321–329. [CrossRef] [PubMed]
8. Galvez, J.; Galvez-Llompert, M.; Garcia-Domenech, R. Application of molecular topology for the prediction of the reaction times and yields under solvent-free conditions. *Green Chem.* **2010**, *12*, 1056–1061. [CrossRef]
9. Blay, V.; Gullon-Soletto, J.; Galvez-Llompert, M.; Galvez, J.; Garcia-Domenech, R. Biodegradability Prediction of Fragrant Molecules by Molecular Topology. *ACS Sustain. Chem. Eng.* **2016**, *4*, 4224–4231.
10. Smlib v2.0. Available online: <http://melolab.org/smlib/> (accessed on 16 February 2019).
11. ChemT. Available online: <http://www.esa.ipb.pt/biochemcore/index.php/ds/c> (accessed on 16 February 2019).
12. Library Synthesizer. Available online: <https://tripod.nih.gov/?p=370> (accessed on 16 February 2019).
13. Virtual Combinatorial Library of cheminfo.org. Available online: http://www.cheminfo.org/Chemistry/Cheminformatics/Virtual_combinatorial_library/index.html (accessed on 16 February 2019).
14. Combinatorial Library Design of e-LEA3D. Available online: <http://chemoinfo.ipmc.cnrs.fr/eDESIGN/reagent.html> (accessed on 16 February 2019).
15. DesMol2 Software. Molecular Connectivity and Drug Design Research Unit. Department of Physical Chemistry. Faculty of Pharmacy. University of Valencia. Available online: <http://desmol2.uv.es> (accessed on 16 February 2019).
16. Besalú, E. Fast Modeling of Binding Affinities by Means of Superposing Significant Interaction Rules (SSIR) Method. *Int. J. Mol. Sci.* **2016**, *17*, 827. [CrossRef] [PubMed]
17. Galvez, J.; Garcia-Domenech, R.; de Gregorio Alapont, C.; de Julian-Ortiz, J.; Popa, L. Pharmacological distribution diagrams: A tool for de novo drug design. *J. Mol. Graph.* **1996**, *14*, 272–276. [CrossRef]
18. *Dragon for Windows (Software for Molecular Descriptor Calculations)*, version 5.4; Talete srl: Milan, Italy, 2006.
19. Klein, D.J.; Lukovits, I.; Gutman, I. On the definition of the hyper-Wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 50–52. [CrossRef]
20. Kier, L.B.; Hall, L.H.; Frazer, J.W. An index of electrotopological state for atoms in molecules. *J. Math. Chem.* **1991**, *7*, 229–241. [CrossRef]
21. Gramatica, P.; Corradi, M.; Consonni, V. Modelling and Prediction of Soil Sorption Coefficients of Non-ionic Organic Pesticides by Molecular Descriptors. *Chemosphere* **2000**, *41*, 763–777. [CrossRef]
22. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors, Methods and Principles in Medicinal Chemistry Series*; WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2009; Volume 41, pp. 719–721.
23. Galvez, J.; Garcia-Domenech, R.; Salabert, M.T.; Soler, R. Charge indexes. New topological descriptors. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 520–525. [CrossRef]
24. PaDEL-Descriptor. Available online: <http://www.yapcsoft.com/dd/padeldescriptor/> (accessed on 16 February 2019).
25. Kier, L.B.; Hall, L.H. *Molecular Connectivity in Structure-Activity Analysis*; Research Studies Press: Letchworth, UK, 1986.

26. Galvez, J.; Garcia-Domenech, R.; de Julian-Ortiz, J.V.; Soler, R. Topological Approach to Drug Design. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 272–284. [[CrossRef](#)] [[PubMed](#)]
27. Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20. [[CrossRef](#)] [[PubMed](#)]
28. StatSoft. *Statistica (Data Analysis Software System)*, version 9; StatSoft: Tulsa, OK, USA, 2009.
29. Roy, P.P.; Leonard, J.T.; Roy, K. Exploring the impact of size of training sets for the development of predictive QSAR models. *Chemom. Intell. Lab. Syst.* **2008**, *90*, 31–42. [[CrossRef](#)]
30. Wold, S.; Eriksson, L.; Clementi, S. Statistical validation of QSAR results. In *Chemometric Methods in Molecular Design*; Wiley-VCH Verlag GmbH: Weinheim, Germany, 1995; pp. 309–338.



© 2019 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 (<http://creativecommons.org/licenses/by/4.0/>).