Revue Roumaine de Chimie, **2007**, *52*(8–9), 789–793

Dedicated to Professor Victor-Emanuel Sahini on the occasion of his 80th anniversary

THE HIERARCHY OF THE MOLECULAR DESCRIPTORS AND THE LIGAND – RECEPTOR INTERACTION

PAUL GABRIEL ANOAICA,[a*] EMILIA AMZOIU [a] and COSTINEL I. LEPĂDATU[b*]

[a] Faculty of Pharmacy, University of Medicine and Pharmacy Craiova, 1 Mai nr. 66, cod 200638, Roumania
[b] Institute of Physical Chemistry "Ilie Murgulescu", Roumanian Academy, Sp. Independenței 202, 060021, Bucharest, Roumania

Received February 1, 2007

In QSAR and CADD procedures, the interface between chemical structures and their biological response is assured by a very large number of descriptors reflecting different aspects or hypostases of the chemical substances involved in the interaction with a biological receptor. A thorough analysis of these descriptors reveals a surprising connection which might exist between certain descriptors labelled as "internal" descriptors directly responsible for the chemical bond formation inside the molecule and the so-called "external" descriptors directly involved in the ligand – receptor interaction. Such a hierarchy can help us to understand the biological activity as well as its optimization.

INTRODUCTION

As we know, QSAR and CADD procedures reported in the literature[1-3] are generally based on drug (ligand) - biological receptor interaction conceived as a chain

{ Structure } $\Leftrightarrow$ **{ Descriptors }** $\Rightarrow$ **{ Biological Activity }**

defined for a class of chemical substances.

The efforts of almost all QSAR studies are centered on deriving linear regression relationships between these descriptors and biological activity, according to the Hansch equation:

$$A = a_0 + a_1X_1 + a_2X_2 + \dots , \qquad (1)$$

where a_0, a_1, ... are regression coefficients, A the biological activity and X_1, X_2, ... structural descriptors respectively. The descriptors involved in equation (1) are of different nature: geometric, topological, thermodynamic, electrostatic, physicochemical and quantomolecular[4].

Despite the apparent simplicity of these principles, the obtained regression equations are not always useful, their predicting power being in many cases extremely poor. The deficiency is primarily due to the huge number of descriptors necessary to describe both the chemical structures from the class of substances and to the lack of information regarding the ligand - receptor interaction.[5] This handicap is reflected in the effort to increase the correlation coefficient value R^2 of the linear models by introducing in the Hansch equation (1) 4-8 (and even more) descriptors. As the number of descriptors available increased, methods are required to select the appropriate descriptors from a large pool of descriptors. Unfortunately, the predictability of the linear regressions containing more than 6 descriptors is reduced only to the studied class of substances.

In these last years, there is a prevailing tendency in QSAR/QSPR research to rationalize the used procedures and quantities (descriptors). Gasteiger[6] suggests a hierarchy of structure representation by means of structure coding methods, the classification and selection of descriptors,[7] or even a hierarchic system of QSAR

* Corresponding author: gabriel@umfcv.ro

models (1D – 4D) on the base of simplex representation of molecular structure.[8]

There are also reported some attempts[9-12] to select a fixed collection of few thought – out descriptors in QSAR by means of which the molecular diversity can in principle be described (chemistry space). It must be mentioned in this context, that only few programs dedicated to QSAR research make a careful selection of the descriptors used[13-14] or they are based on the principles regarding the ligand – receptor interaction.[15]

The aim of this paper is to suggest a new way to limit the number of descriptors involved in QSAR based on the assumed hierarchy which must exist between descriptors as results from the analysis of the descriptors according to their designation and meaning. As we shall see in the following, a possible hierarchy or interdependence between descriptors can be observed if the linear regression equations of different orders are compared.

The descriptors involved in equation (1) can in this way be categorized into **internal** descriptors having to do with the chemical bond formation in each molecule and **external** descriptors, directly bound with the ligand – receptor interaction.

This classification can help us to identify the mentioned thought – out descriptors by means of which the biological response for a set of substances can be characterized.

The hierarchy between descriptors enounced in this paper might also help us to understand the way how different substituents via the internal descriptors can influence the ligand - receptor interactions. We believe that the internal and external descriptors are in principle interdependent and a thorough analysis of the nature of these quantities could be helpful for QSAR research.

RESULTS AND DISCUSSION

We shall present in the following, two QSAR studies illustrating the supposed hierarchy or interdependence between descriptors involved in the structure – activity relationships. Further studies have also confirmed this assumption.[16]

The antimycobacterial activity for a series of 2-halogeno-6-alkylsulfanyl-4-amidopyridines (**1**-**8**), 2-Alkylsulfanyl-6-hexylsulfanyl- 4-amidopyridines (**9**-**13**) and 2-Halogeno-6-alkylsulfanyl-4-thioamidopyridines (**14**-**21**) derivatives are given in Figure 1 and Table 1. The antimycotic and antibacterial activity experimentally determined for these compounds have been reported in literature[17]. We have respected in Table 1 the original labeling for the compounds given by the cited authors.

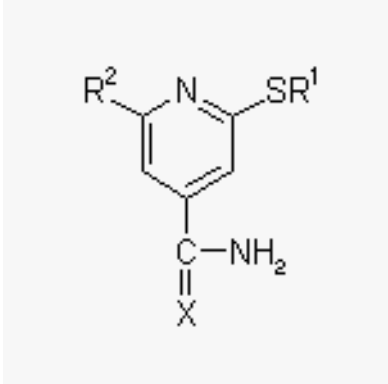


Fig. 1 – Structures of amidopyridine and thioamidopryridine derivatives.

Table 1

The antimycobacterial activity of 21 amidopyridines and thioamidopyridines

Compound	R^1	R^2	X	IC50* (μmol/l)	Compound	R^1	R^2	X	IC50* (μmol/l)
1	C_2H_5	Cl	O	101.5	12	C_6H_{13}	SC_7H_{15}	O	543.6
2	C_3H_7	Cl	O	58.4	13	C_6H_{13}	SC_8H_{17}	O	258.8
3	C_6H_{13}	Cl	O	10.2	14	C_2H_5	Cl	S	104.8
4	CH_3	Br	O	76.7	15	C_3H_7	Cl	S	9.3
5	C_2H_5	Br	O	34.2	16	C_6H_{13}	Cl	S	29.8
6	C_4H_9	Br	O	10.6	17	CH_3	Br	S	187.7
7	C_7H_{15}	Br	O	5.9	18	C_2H_5	Br	S	19.6
9	C_6H_{13}	SC_2H_5	O	9.1	19	C_4H_9	Br	S	20.9
10	C_6H_{13}	SC_4H_9	O	203.5	20	C_7H_{15}	Br	S	61.0
11	C_6H_{13}	SC_5H_{11}	O	249.3	21	C_8H_{17}	Br	S	105.1

* -logIC50(mole/l) values have been used in the regression analysis.

The molecular structures given in Table 1 have been modeled using standard procedures (Molecular Mechanics: MM+ followed by the geometry optimization using MOPAC 7: RHF, PM3)[18]. The molecular features of the derivatives were calculated from MOPAC outputs using the CODESSA 2.64 software.[19] The last program has been preferred because it offers a large variety of descriptors (constitutional, topological, geometrical, electrostatic, quantum-chemical, MO-related and thermodynamic).[4,19] The obtained descriptors have been correlated with –log(IC50) using multiple regression analysis -log (IC50) = $a_0 + \sum a_i X_i$, where –log IC50 is the response, and X_i are the descriptors, a_0, a_1, a_2, …the regression coefficients. Table 2 summarizes the results of the statistical analysis for i = 4 ÷ 2, the descriptor order being given upon the decreasing values of the a_i coefficients:

Table 2

The multiliniare regression analysis –log(IC50)=$a_0+a_1X_1+a_2X_2+\ldots$, X_i= descriptors

4 descriptors: best correlations	descriptors involved:
1: R2=0.9670 (4 descriptors) 266 267 275 45	266 - Max coulombic interaction for a C-H bond
2: R2=0.9632 (4 descriptors) 266 267 45 279	267 - Min total interaction for a C-H bond
3: R2=0.9629 (4 descriptors) 237 267 45 121	275 - Min electron-nuclear attraction for a C-S bond
4: R2=0.9626 (4 descriptors) 266 267 45 281	45 - Max atomic nucleoph. react. index for a C atom
5: R2=0.9625 (4 descriptors) 237 255 110 111	279 - Min coulombic interaction for a C-S bond
6: R2=0.9621 (4 descriptors) 266 267 86 45	237 - Min coulombic interaction for a C-C bond
7: R2=0.9620 (4 descriptors) 266 267 275 23	**121 - RNCS**
8: R2=0.9619 (4 descriptors) 266 267 45 273	281 - Min total interaction for a C-S bond
9: R2=0.9605 (4 descriptors) 237 267 45 323	255 - Min resonance energy for a C-H bond
10: R2=0.9584 (4 descriptors) 266 267 86 23	**110 - WNSA-2**
	111 - PPSA-3
	86 - Min net atomic charge for a C atom
	23 - Final heat of formation
	273 - Min electron-electron repulsion for a C-S bond
	323 - Principal moment of inertia B
3 descriptors: best correlations:	descriptors involved:
1: R2=1.0000 (3 descriptors) 28 159 210	**28 - Average atomic nucleoph. react. index for a O atom**
2: R2=1.0000 (3 descriptors) 28 159 51	159 - Min coulombic interaction for a C-O bond
3: R2=1.0000 (3 descriptors) 202 189 102	**210 - 1X GAMMA polarizability (DIP)**
4: R2=1.0000 (3 descriptors) 202 189 196	**51 - Total dipole of the molecule**
5: R2=1.0000 (3 descriptors) 28 159 78	**202 - Principal moment of inertia A / # of atoms**
6: R2=1.0000 (3 descriptors) 202 189 101	189 - Min total interaction for a C-H bond
7: R2=1.0000 (3 descriptors) 202 189 198	102 - Min total bond order (>0.1) of a C atom
8: R2=1.0000 (3 descriptors) 28 159 49	196 - Total molecular two-center resonance energy / # of atoms
9: R2=1.0000 (3 descriptors) 202 189 60	**78 - RPCS**
	101 - Average valency of a C atom
	198 - Total molecular two-center exchange energy / # of atoms
	49 - Total point-charge component of the molecular dipole
	60 - WPSA-1
2 descriptors: best correlations :	descriptors involved:
1: R2=0.8855 (2 descriptors) 237 255	237 - Min coulombic interaction for a C-C bond
2: R2=0.8807 (2 descriptors) 237 267	255 - Min resonance energy for a C-H bond
3: R2=0.8728 (2 descriptors) 266 26	267 - Min total interaction for a C-H bond
4: R2=0.8692 (2 descriptors) 266 255	266 - Max coulombic interaction for a C-H bond
5: R2=0.8542 (2 descriptors) 237 175	175 - Max total bond order of a H atom
6: R2=0.8538 (2 descriptors) 235 255	235 - Min nuclear-nuclear repulsion for a C-C bond
7: R2=0.8405 (2 descriptors) 235 267	265 - Min coulombic interaction for a C-H bond
8: R2=0.8294 (2 descriptors) 265 229	229 - Min exchange energy for a C-C bond
	26 - No. of occupied electronic levels / # of atoms

As may be seen, the descriptors occurring in the linear regressions of i = 2,3,4 order have quite different provenance: quantum – chemical, MO-related, geometrical, electrostatic and thermodynamic. All these descriptors selected by the heuristic regression procedure can be classified into: 1 – **internal** descriptors, directly involved in the "internal" chemical bond formation of each molecule, such as Coulomb interaction for a C-C bond (237), minimal resonance energy for a C-H

bond (255), etc. 2 – **external** descriptors, directly bound with the ligand – receptor interaction, such as molecular descriptors (moments of inertia 323, 202) or the descriptors describing the partition of the positive or negative charges on the molecular surface area (121, 110, 111, 78, 60) or even a nucleophilic reaction index (28). These descriptors called as external are shown in Table 2 with bold letters.

Descriptors 323, 202 have undoubtedly to do with the molecular shape and the descriptors 121 (RNCS – Relative Negative Charge Surface), 110 (WNSA – Weighted Negative (charged) Surface Area), 111 (PPSA – Partial Positively (charged) Surface Area), 78 (RPCS – Relatively Positive Charged Surface) and 60 (PSA – Weighted Positive (charged) Surface Area) with the first step of molecular recognition based on weak electrostatic ligand – receptor interactions. The significance of these descriptors is described in the literature.[4,19-22]

As far as the coefficients a_i in the statistical linear equation (1) is concerned, a comparative analysis could be done in the following situations:

In the frame of the same equation estimating the contribution of each descriptor to the formation of the biological answer upon the values taken by a_i coefficients.

For the equations having the same number of descriptors one may signalize the presence or the absence of certain descriptors in the studied equations. The descriptors having a lower weight (a_i coefficient value) will obviously not appear as the number of descriptors in the statistical equations becomes smaller.

The present paper does not follow the history of the descriptors, but to choose the best statistical equations having R^2 correlation coefficients close to unit. We get in this way a data basis which makes possible the classification or hierarchy of the descriptors involved upon their significance and nature.

As may be seen in Table 2, the internal descriptors are present in all statistical equations as a background, even though their appearance is not entirely justified by the physical and chemical principles governing the ligand – receptor interaction.

This fact could be explained by the connection existing between the two categories of internal and external descriptors, even though such a connection cannot quantitatively be done. As an example, the descriptors for the molecular shape (external) depend on the space arrangement of the atoms in molecule, ultimately due by a series of factors (internal descriptors) such as the nature of the atoms, a multitude of factors conditioning the chemical bond formation, stabilization energy of the molecular system, etc.

The electrostatic descriptors involved in the example herein discussed, depend on the partition of the electron population in the molecule, conditioned in its turn by the multitude of internal descriptors including those previously enumerated. We could in this way explain the presence of many internal descriptors in the statistical equations, even though they have not directly involved in the ligand – receptor interaction.

As may be seen in Table 2, on going from i = 4 to i = 3, 2, in other words, if the number of descriptors in the regression process is gradually decreased, one may observe that only "internal" molecular descriptors remain, i.e. those descriptors intimately connected with formation of the chemical bonds inside the molecule. The apparently disappearance of the external descriptors from the regression equation for lower "i" order could actually be interpreted as the connection or interdependence which must exist between the internal and external descriptors. As an example, the descriptors regarding the maps of positive / negative electric charges of the atoms in molecule spread out on the molecular surface area (RNCS, WNSA, etc.) are undoubtedly related to the internal quantum – chemical quantities responsible for the partition of the electron population.

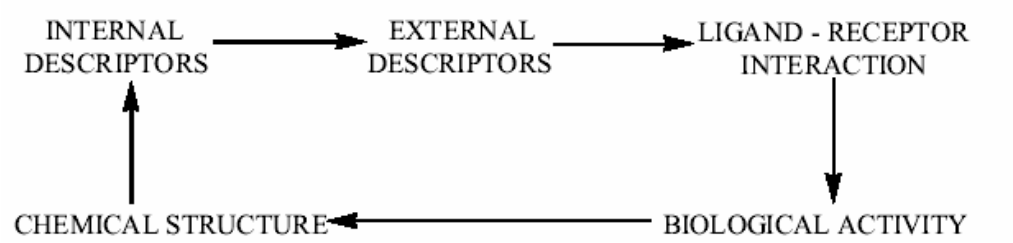
The hierarchy of the descriptors occurring in regression analysis of the chain Structure ↔ Descriptors → Biological Response can help us to understand the influence of the different chemical groups on the structural descriptors and their influence on the biological activity.

The identification of these internal descriptors can represent a first step toward the selection of thought - out descriptors, because they must contain information about the changes caused by presence or by the nature of different chemical groups inside the molecule.

The QSAR analysis performed for other classes of substances exerting a specific biological activity have revealed the same hierarchy between descriptors categorized as internal and external.

CONCLUSION

We can imagine a method, according to the following scheme going backwards from biological activity to the structure.



One may analyze the connection which should exist between the external and internal descriptors and the influence of different chemical groups on the biological activity. A better understanding of the nature of the structural descriptors and their hierarchyzation can open a way to limit the number of descriptors involved in the ligand – receptor interactions offering further information about the mechanisms of these interactions.

REFERENCES

1. C. Hansch and A. Lee, "Exploring QSAR: Fundamentals and Applications in Chemistry and Biology", American Chemical Society, 1995.
2. K.F. Koehler, S.N. Rao and J.P. Snyder, "Modeling Drug – Receptor Interactions" in "Guidebook on Molecular Modeling in Drug Design" (Ed.) N. Claude Cohen, Acad. Press, New York, 1995.
3. Z. Simon and T.I. Oprea, "QSAR – an overview receptor – effector interaction" in "Relatii Cantitative Structura Chimica – Activitate Biologica", Grupul QSAR Timişoara, Editura Mirton, 1996.
4. A.R. Katritzky, R. Petrukhin, H. Yang and M. Karelson, "Comprehensive Descriptors For Structural and Statistical Analysis Codessa Pro (reference manual)", University of Florida, 2003.
5. M.T.D. Cronin and T. Wayne Schultz, *J. Mol. Structure (Theochem)*, 2003, *39*, 622.
6. J. Gasteiger " A Hierarchy of Structure Representations" in "Handbook of Chemoinformatics - From Data to Knowledge" (Ed.) J. Gasteiger, Wiley-VCH, Weinheim, 2003.
7. C.M. Breneman, C.M. Sundling, N. Sukumar[1], L. Shen, W.P. Katt. and M. J. Embrechts, *J. Comp.-Aided Mol. Design*, 2003, *17,*231.
8. V.E. Kuzmin, A.G. Artemenko, P.G. Polischuk, E.N. Muratov, A.I. Hromov, A.V. Liahovskiy, S.A. Andronati and Svetlana Y. Makan, *J. Mol. Model.*, 2005, *11*, 457.
9. P. Labute, J. Molec. Graph. & Modelling, 2000, *18*, 464.
10. T. Clark, *J.Mol.Graphics Model*, 2004, *22*, 519.
11. M. Brustle, B. Beck, T. Schindler, W. King, T. Mitchell, and T Clark, *J. Med. Chem.*, 2002, *45*, 3345.
12. M. Olah, C. Bologa and T.I. Oprea, *J. Comp.- Aided Mol. Design*, 2004, *18*, 437.
13. L. Tarko, O. Ivanciuc, QSAR Modeling of the Anticonvulsant Activity of Phylacetanilides with PRECLAV (Property Evaluation by Class Variables), *Commun. Math. Chem.* (MATCH), 2001, *44*, 201.
14. L. Tarko, I. Lupescu, D. C. Groposila, Sweetness power QSARs by PRECLAV software, *Arkivoc*, 2005.
15. Z. Simon, A. Chiriac, S. Holban, D. Ciubotariu and G. I. Mihalas: "Minimum Steric Difference. The MTD Method for QSAR Studies", Res. Stud. Press. Ltd. (John Wiley), Letchworth, England, 1984.
16. P.G. Anoaica, "Optimizarea activității prodrug-urilor prin metode CADD", PhD Thesis, 2007.
17. M. Miletin, J. Hartl, M. Dolezal, Z. Odlerova, K.Kralova and M. Machacek, *Molec.*, 2000, *5*, 208.
18. MOPAC 7.0 for UNIX, Quantum Chemistry Program Exchange, Project 688.
19. A. R. Katritzky, M. Karelson and R. Petrukhin, "Comprehensive Descriptors for Structural and Statistical Analysis", University of Florida, Gainsville (Semichem).
20. M. Karelson, "Molecular Descriptors in QSAR/QSPR, J. Wiley & Sons, New York, 2000.
21. D.T. Stanton, P.C. Jurs, *Anal. Chem.*, 1990**,** *62*, 2323.
22. D.T. Stanton, L.M. Egolf, P.C. Jurs, M.G. Hicks, *J. Chem. Inf. Comput. Sci.*, 1992**,** *32*, 306.