QSAR and QSPR study of derivatives 4-arylaminocoumarin

Davorka Završnik, Samija Muratović, Selma Špirtović
University of Sarajevo, Faculty of Pharmacy, Department of Pharmaceutical chemistry Čekaluša 90, Sarajevo

Abstract

Coumarin and its derivatives are reactive compounds, suitable for many syntheses. They are used as anticoagulants, antibacterial, animistic compounds. The interest in coumarins has increased because it was found that they reduce the HIV virus activity.
The synthesis of 4-arylaminocoumarin derivatives from 4-hydroxycoumarin, has been carried out, and their antimycotic effects were tested. In the QSAR (quantitative structure-activity relationship) QSPR (quantitative structure-property-activity relationship) study we have used physicochemical properties and topological indices (Balaban index J(G), Wiener index W(G), information-theoretical index I(G), and valence connectivity index (G), to predict bioactivity of the newly synthesized coumarin compounds. By using methods of molecular modelling, the relationships between structure, properties and activity of coumarin compounds have been investigated.
The best QSPR models were obtained using valence connectivity index or combination indices. According Rekker's method the best correlation of calculated values log P, has been obtained with the model based on the inhibition zone (I) 4-arylaminocoumarin derivatives expressed in mm.
The results obtained in this study enable further synthesis of new coumarin derivatives and predict their biological activity and properties.

Keywords: QSPR, QSAR, derivates of 4-arylaminocoumarin

Introduction

Studies of natural and synthetic coumarins and its derivatives have been present for a number of years. Coumarins and their derivatives are characterised with very good chemical reactivity and different bioactivity. A great number of synthesized derivatives are biologically active, and many of them are applied in therapy as anticoagulant, antibacterial and antifungal agents (1). Recently the interest in coumarins has increased significantly because it was found that they reduce the HIV virus activity. (2, 3) A molecule of 4-hydroxycoumarin is very reactive and suitable for many syntheses. Due to the exceptional reactivity, we synthesized new 4-arylaminocoumarin derivatives from 4- hydroxicoumarin
Derivatives of 4-arylaminocoumarin from 4-hydroxicoumrins were synthesized and their microbiological activity were examined. Newly prepared coumarin derivatives have various constituents, and according to that, they can exhibit potential microbiological activity; therefore, the microbiological activity of these derivatives in case of various species of bacteria and fungi was tested. The results obtained indicate that newly synthesized compounds have much better animistic properties than antibacterial (4).
QSAR analysis is a useful tool for examining the relationship between the biological activities, the physicochemical properties and the molecular structures of a series of compounds. The fundamental axiom of QSAR and QSPR modelling is that the structures of molecules are reflected in their biological activities and physicochemical properties (5). Most molecular activities and properties can be represented by a single number, but molecular structure cannot similarly be represented. The representation of molecular structures by numbers is a way to encode the structural information in QSAR and QSPR studies. The modelling process reduced to a correlation between two set of numbers, one sets of numbers representing the molecular bioactivity or property and the other set representing the molecular structure. This correlation is meaningful only if it is carried out for a larger set of molecules.
In this work, several quantitative QSPR models and QSAR models with topological indices will be used to study of some physicochemical properties and bioactivity of newly synthesized coumarins have been investigated. Topological indices have been mainly used for purpose of correlating the properties of molecules with their topological structure, since most of biological processes are difficult to quantify accurately. Lipophilicity has been investigated by determining the log P (partition coefficient). Lipophilicity is a property of a molecule, which depends on structure and can be changed by the modifications in the molecular structure. (6)

Material and methods

Synthesis of derivatives of 4-arylaminocoumarin and their microbiological investigation

The synthesis of 4-arylaminocoumarins derivatives and their spectral characteristics, elementary analysis and results of microbiological activity were described in our previous investigation (4). The microbiological activity

of compounds was tested by the diffusion method on species Candida albicans 5934 (4) and are shown as inhibition zones expressed in mm. The results of synthesis, structure of 4-arylaminocoumarin derivatives and their microbiological activity are shown in Table 1.

According to the method of encoding the structural information, the QSAR and QSPR models may be classified into three major groups (7):

1. empirical models
2. quantum - chemical models
3. non-empirical models.

In our research, we applied non- empirical QSAR-QSPR models, topological indices and log P.
In QSAR modelling of lipophilicity of our compounds, the following topological indices were used: the Wiener index. The valence-connectivity indices, the Balaban index and information-theoretical index. In order to describing of molecular properties our compounds we are used values of log P. These values for the studied compounds were calculated according to the method of Rekker (8).

Topological indices

The valence connectivity index, χ^v (G)

The molecular connectivity index, $\chi^v(G)$ of molecular graph G is defined as (9):

$$\chi(G)=\sum[d(i)d(j)]^{-0.5}$$

where the sum is taken over all edges of G. In equation d (i) and d (j) are the valences of vertices i and j making up edge i-j.
For heterosystems, the connectivity index is given in terms of valence delta values δ (i) and δ (j) of atoms i and j. In that case it is denoted by χ^v.
This type of connectivity index is called the valence connectivity index (10) and is defined as

$$\chi^v=\sum_{ij}[\delta(i)\,\delta(j)]^{-0.5}$$

where the sum is taken over all bonds i-j of the molecule. In this case edge of G has a weight of δ(i) δ(j).

The Balaban Index, J (G)

Balaban index J (G) represents the extended connectivity. This index, denoted by J (G), is defined as (11):

$$J(G) = (M/\mu+1)\sum_{edges}(didj)^{-0.5}$$

Where is:

M- the number of edges in G;
μ- cyclomatic number of G ;
di- distance sum (i =1, 2, 3,…., N-Number of vertices in G).

The distance sum for a vertex i represents the sum of all entries in the corresponding row (or column) of the D distance matrix:

$$d_i = \sum_{j=1}^{N}(D)_{ij}$$

The cyclomatic number μ of a polycyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it onto the related acyclic subgraph:
$\mu = M-N+1$

Barysz at al. have modified the elements of the distance matrix for heterosystems (12):

1) diagonal elements
$(D)_{ij} = 1-(Z_c/Z_i)$
$Z_c = 6$ and Z_i is the atomic number of the given element;

2) off-diagonal elements

$$(D)ij = \sum_r k_r$$

where k_r, the bond parameter is:

$k_r = (1/b_r)\,(Z_c^2/ Z_iZ_j)$

b_r is the bond weight that has values 1 for single bond, 2 for double bond. For different types of heterobonds, the values are given in the literature (13).

The Information-Theoretic Index, I(G)

This index was calculated by the modification of Shannon´s relation (14):

$$I(G) = -\sum_{i=1}^{n}\frac{2N_1}{N(N-1)}\log\frac{2N_1}{N(N-1)}$$

n-the number of different sets of elements
N_i-the number of elements in the i-th set of elements and the sum is over all sets of elements.

The Wiener Index, W(G)

The Wiener Index, W(G) of a structure G can be obtained from the distance matrix D of the corresponding hydrogen-depleted chemical graph G as the half-summation of the elements of D(15):

$$W(G) = 1/2 \sum_{i,j} (D)_{ij}$$

$(D)_{ij}$- off-elements of D(G) which stand for the shortest distance in terms of the number of bonds between atoms i and j in G.

The calculation of partition coefficient logP (octanol/water)

Values the log P(o/w) for the series coumarines are calculated according to the method
of Nys and Rekker (16,17):

$$P(o/w) = \sum_{i=1}^{n} a_n f_n$$

f- the hydrophobic fragmental constant, the lipophilicity contribution of a constituent part of a structure to the total lipophilicity.

a_n- numerical factor which indicating the incidence of fragment (f_n) in the structure.

TAM program for the calculation of topological indices in QSPR and QSAR studies was used.

Results

The results of synthesis, structure of 4-arylaminocoumarin derivatives and their microbiological activity and selected physiochemical properties: molecular mass (M.m.), melting point (M.p.) partition coefficient (logP) calculed according to Rekker's method, van der Waals-volume (Vw) and inhibition zones of arilaminocoumarin derivatives expressed in mm, and topologic indices: Wiener index W(G), information-theoretical index I(G), Balaban index J(G) and valence connectivity index $\chi^v(G)$ are shown in Table 1.

Discussion

In this work, several quantitative QSPR and QSAR models for predicting properties and activity of 4-arylaminocoumarin derivatives are considered. As one might expect, different properties were best modelled with different regression and different indices. We have compared each physicochemical property with each topological index. To test the quality and accuracy of derived models, the following statistical parameters were used: **n** is the number of data points, **r** the correlation coefficient, **s** the standard deviation.
Molecular mass shows best correlation with valence connectivity index $\chi^v(G)$ (Figure 1). It is obvious that by increase of the molecule by substituents CH_3, OCH_3 and OC_2H_5 the correlation is decreased.

Figure 1 The parabolic QSAR model of molecular masses (Mm) of arylaminocoumarin derivatives, based on valence connectivity index $\chi^v(G)$

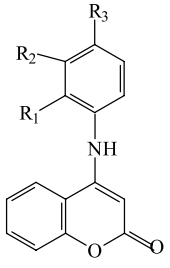
$\chi(G)=50.1754 - 0.3544\ Mm + 0{,}0007\ Mm^2$
n=10
r= 0.9039
s= 0.14

Van der Walls volume shows best correlation with valence connectivity index $\chi^v(G)$

Figure 2 The linear QSPR model of van der Walls volume v(W) of 4-arylaminocoumarin derivatives based on the valence connectivity index $\chi^v(G)$

$\chi^v(G)= 3.2435 + 1.3225\ V(w)$
n=10
r= 0.9667
s= 0.08

Table 1. The structure of prepared 4-arylaminocumarins, the selected physiochemical properties: molecular mass (M.m.), melting point (M.p.) partition coefficient (logP) calculed according to Rekker's method, van der Waals-volume (Vw) and inhibion zones of arilaminocoumarin derivatives expressed in mm, and topologic indices: Wiener index W(G), information-theoretical index I(G), Balaban index J(G) and valence connectivity index $\chi^v(G)$.

N°	R_1	R_2	R_3		Molecular formula	Mm	Tt	W(G)	I(G)	J(G)	^{v}(G)	Vw	log P
1	H	H	H		$C_{15}H_{11}O_2N$	237	267-8	755	3.293	2.2228	5.9314	2.017	-0.905
2	OCH_3	H	H		$C_{16}H_{13}O_3N$	267	223-5	1007	3.352	2.1630	6.1540	2.111	-2.643
3	H	H	OCH_3		$C_{16}H_{13}O_3N$	267	245-6	1055	3.509	2.0559	6.1480	2,111	-2.643
4	CH_3	H	H		$C_{16}H_{13}O_2N$	251	214-6	877	3.309	2.1174	6.0416	2.171	-0.401
5	H	CH_3	H		$C_{16}H_{13}O_2N$	251	212-3	889	3.343	2.1117	6.0357	2.171	-0.401
6	H	H	CH_3		$C_{16}H_{13}O_2N$	251	279-280	901	3.401	2.0827	6.0357	2.171	-0.401
7	OC_2H_5	H	H		$C_{17}H_{15}O_3N$	281	162-4	1157	3.430	2.1377	6.7415	2.630	-2.113
8	H	H	OC_2H_5		$C_{17}H_{15}O_3N$	281	218-220	1229	3.612	2.0023	6.7356	2.630	-2.113
9	H	Cl	H		$C_{15}H_{10}O_2NCl$	271.5	202-5	973	3.487	2.4824	6.0530	2.182	-0.097
10	H	H	Cl		$C_{15}H_{10}O_2NCl$	271.5	306-7	901	3.401	2.2554	6.1376	2.182	-0.097

The valence connectivity index $\chi^{v}(G)$ shows best correlation with Wiener index W(G)(Figure 3).
Connectivity index characterizes the structure best in these investigations. The sizes of the molecule and ramification are characterised by the valence connectivity index, the compounds are much related and characteristic groups methyl, methoxy, contributed the increase valence connectivity index as well as the Wiener index.

Figure 3 The parabolic QSPR model of valence connectivity index $\chi^{v}(G)$ of 4-arylaminocoumarin derivatives based on the Wiener index (W(G).

$W(G)= 12771{,}32\ \chi^{v}(G) - 40988.60 - 966.31\ \chi^{v}(G)^{2}$
n=10
r= 0.9482
s= 51.16

Table 1 shows the values of log P which for the investigated compounds are all negative, indicating a decrease of lipophilicity. The QSAR model based on the partition coefficient has shown the best correlation with the inhibition zone (I) 4-arylaminocoumarin derivatives expressed in mm (Figure 4).

Figure 4 The parabolic QSPR model of inhibition zone I (mm) of 4-arylaminocoumarin derivatives based on the partition coefficient log P

I =19.8574 + 9.1083 log P + 1.8379 log P2
n=10
r= 0.9121
s= 2.28

The partition coefficient was found most suitable in QSPR study and correlation with tested activity to Candida albicans 5934 is shown as the inhibition zone (I) of certain 4-arylaminocoumarin derivatives expressed in mm. The best activity to Candida albicans was shown by the compounds 4 and 9. Better activity to Candida albicans was shown by the compound with substituent on R1 position. The results obtained indicate that zone inhibition increases with lipophilicity.

Conclusion

We have investigated linear and several nonlinear relationships between the topological indices as discussed in previous section, and selected properties of 4-arylaminocoumarin derivatives.

In our investigations, the best QSPR models were obtained using valence connectivity index or combination indices. The best QSAR model was based on the partition coefficient and it has shown the best correlation with the inhibition zone (I) 4-arylaminocoumarin derivatives expressed in mm. All newly synthesized compounds have negative values of log P, and indicate decrease of lipophilicity. CH_3 group and Cl increased lipophilicity and antimycotic activity, especially when the substituent was on R1 position. Other examined indices and regression analysis have not a good correlation and are irrelevant for the purpose of this study.
The results obtained in this study will enable targeted synthesis of new coumarin derivatives of defined physicochemical properties and desired biological activity.

References

1. Mutschler E., Schafer-Korting M. Arzneimittel-Wirkungen. Wissenschaffliche Verlagsgesellschafft mbH Stuttgart 1997; 427- 429

2. Kampranis S.C.,Gormley N. A., Tranter R.,Orphanides G., Maxwell A. Probing the binding of Coumarins and Cyclothialidines to DNA Gyrase. Byochemistry 1999; 16: 38 (7): 1967- 1976

3. Zhao H., Neamati N., Hong H .X., Mazumder A.,Wang S. M., Sunder S., Milne G.W. A. Coumarin Based Inhibitors of HIV Integrase. J.Med. Chem. 1997; 40: 242-249

4. Završnik D., Gaon D. Sinteza, struktura i mikrobiloška Aktivnost 4- arylaminokumarina. Pharmacia 1999; 10 (2): 28-33

5. Trinajstić N. Chemical graph Theory, 2nd ed., CRC Press, Boca Raton, F.L., 1992.

6. Nys G.G., Rekker R.F. Clinica Therapeutica. Eur. J. Med. Chem.1974; 9: 361

7. Trinajstić N. Topological indices in: Chemical Graph Theory, Klein, D.J. and Randić, M., (Eds), 1992, pp. 225-273, Florida, Boca Raton.

8. Rekker R. F. and Mannhold, R. Calculation of Drug Lipophilicity, 1992, Weinheim, VHC Verlagsgesellschaft.

9. Medić-Šarić, M., Maleš, Ž., Šarić, S., Branther, Quantitative Modelling of Flavonoid Glycosides Isolated from Paaliurus spina-christi Mill., Coatica Chemica Acta, 69(4) (1996) 1603-1616.

10. Kier, L. B., Hall, L. H., Molecular Connectivity in Structure-Activity Analysis, Wiley, New York, 1986.

11. Balaban A.T., Chemical Applications of Graph Theory, Academic, London 1976.

12. Barysz, M., Jasari, G., Lall, R.S., Srivastava, V.K. and Trinajstić, N., Chemical Applications of Topology and Graph Theory, King, R. B. (Ed), Elsevier, Amsterdam, 1983, pp.221-230.

13. Rekker R. F. and Mannhold, R. Calculation of Drug Lipophilicity, 1992, Weinheim, VHC Verlagsgesellschaft.

14. Shannon, C. and Weaver, W. Mathematical Theory of Communication, Urbane, University of Illinois, 1949.

15. Rosie A. M. Information and Communication Theory, Van Nostrand Reinhold Co., London, 1973.

16. Medić-Šarić M., Franić D., Debeljak Ž., Lipofilnost lijekova II dio: Rekkerov fragmentni sustav, Farm. Glas. 54 (1998)73-78.

17. Nikolić S. et al. QSPR and QSAR study of phtalimidohydroxamic acids, Acta Pharm, 45(1995)15-24.