QSAR study of synthetic 3-arylcoumarins: *in silico* clastogenic prediction

Estela Guardado Yordi[1,2], Enrique Molina[1,2], Amaury Pérez[3], Lianne León[2], Lourdes Santana,[1] Eugenio Uriarte[1,4] and Maria João Matos[1,*]

[1] Universidad de Santiago de Compostela, Facultad de Farmacia, Campus vida, 15782 Santiago de Compostela, Spain
[2] Universidad de Camagüey Ignacio Agramonte Loynaz, Facultad de Ciencias Aplicadas, Circunvalación Norte Km 5 1/2, Camagüey, Cuba
[3] Universidad Estatal Amazónica, Facultad de Ciencias de la Tierra, Km 2 1/2 vía Puyo a Tena (Paso Lateral), Puyo, Ecuador
[4] Instituto de Química Aplicada, Universidad Autónoma de Chile, Pedro de Valdivia 425, 7500912 Santiago, Chile

* E-mail: mariajoao.correiapinto@usc.es

Abstract. Discovering drugs to a disease is still a challenging task for researchers due to the complexity of biomolecules involved in pathologic processes. Design and development of new and more efficient drugs is still urgent for several diseases. Cheminformatics tools are useful to better understand the complex structures of chemical compounds and the implication of chemical features in the activity. In the current work, a series of synthetic 3-arylcoumarins, with reported antioxidant activity, was studied. A virtual screening, based on the TOPSMODE approach, using a clastogenic model, was performed to predict the potential genotoxicity of the studied molecules. A preliminary interpretation of the relationship between structure and clastogenicity suggests the importance of hydroxyl groups at positions 7 and/or 8 of the coumarin ring. This communication is focused on cheminformatics and its applications on drug discovery, helping to find solutions to complex diseases.

Keywords. 3-Arylcoumarins; Clastogenicity; TOPSMODE; Cheminformatics.

Introduction. The structural diversity of compounds derived from the coumarin scaffold allows to count on molecules that have diverse purposes like food additives, perfumes, cosmetics, agrochemicals and drugs [1-3]. The possibilities of studying new chemical coumarinic structures as potential drugs is a field of great interest because of the synthetic versatility of the scaffold [4]. The coumarin ring has the ability to establish various types of non-covalent interactions including hydrophobic, π-π and electrostatic interactions, as well as hydrogen bonds, van der Waals forces, among others, with active site of several biomolecules, that are important targets of living organisms. The unique feature of containing oxygen in the lactone ring makes these molecules a type of ligand ideal for producing supramolecular assembly.

These molecules have been shown to be effective in a wide range of biological and pharmacological targets. There are numerous reports that recognize their pharmacological interest as anticoagulants, antivirals, enzymatic inhibitors, antibacterial, among others[5]. The antioxidant activity has been extensively studied in the last years, and has been reported for natural and synthetic coumarins. Recently, our research group has experimentally studied the antioxidant activity of several synthetic coumarins [6-11]. This activity is of great interest in the treatment of chronic diseases [12]. In this sense, the modulation

of the coumarin ring, specifically when introducing a heteroaromatic ring (i.e. pyrazole) allows to obtain interesting antioxidant drugs [13].

As said before, series of coumarins, specially 3-arylcoumarins, have been studied and reported as antioxidants by the group [7,11]. Part of this series of coumarins, and some new ones, form the basis of new research that has been based on previous results related to the safety of natural coumarins in food sources and the use of computational toxicology [14,15]. To increase the safety of these molecules as potential drugs, the present work aims to predict clastogenic activity using a cheminformatics (a topological method) previously validated for several families of polyphenolic compounds.

Materials and methods. A virtual screening using the TOPSMODE approach and a QSAR model to predict the clastogenic activity were performed (Equation 1, [16]).

$$\begin{aligned} GT = {} & 0.0091\left[\Omega\left(\mu_1^{PS}\right)\right] - 1.5520\times10^{-4}\left[\Omega\left(\mu_5^{vdW}\right)\right] + 0.148\left[\Omega\left(\mu_4^{Ch}\right)\right] - 0.0021\left[\Omega\left(\mu_2^{PS}\right)\right] + \\ & + 2.6261\times10^{-4}\left[\Omega\left(\mu_3^{PS}\right)\right] - 3.8422\times10^{-5}\left[\Omega\left(\mu_4^{PS}\right)\right] + 1.1520\times10^{-4}\left[\Omega\left(\mu_4^{MR}\right)\right] + \\ & + 1.2011\times10^{-6}\left[\Omega\left(\mu_5^{PS}\right)\right] - 9.8202\times10^{-5}\left[\Omega\left(\mu_5^{MR}\right)\right] - 3.8263\times10^{-5}\left[\Omega\left(\mu_8^{H}\right)\right] - \\ & - 0.0626\left[\Omega\left(\mu_2^{Pol}\right)\right] + 1.6689\left[\Omega\left(\mu_1^{Pol}\right)\right] - 0.0078\left[\Omega\left(\mu_5^{Ch}\right)\right] + 0.1123\left[\Omega\left(\mu_3^{Ch}\right)\right] - 0.6517 \end{aligned} \quad (1)$$

Statisticians: Wilks´- λ= 0.629; F(14.194)=8.148; D^2=2.353; p<0.0000

The Ω is used to indicate that the corresponding variable in brackets was orthogonalized respecting to the rest of the variables included in the model. The classification model obtained is given below, together with the statistical parameters of the linear discriminate of the squared analysis, where λ is the Wilks' statistics, D^2 is the Mahalanobis distance and F is the Fisher ratio.

Molecular descriptors (spectral moments) were weighted to different binding properties: bond distance (SD), standard bond dipole moments (DM), hydrophobicity (H), polar surface area (PS), polarizability (Pol), molar refractivity (MR), van der Waals radii (vdW), and Gasteiger-Marsili charges (Ch). Various software were used to create the database and the model: Chembiodraw Ultra 13.0 for the molecular drawing and obtaining SMILE codes; MODESLAB 1.0 for the calculation of the descriptors; STATISTICA 4.13 for the implementation of linear discriminant analysis for classification in active or inactive. Figure 1 summarizes the procedure carried out.

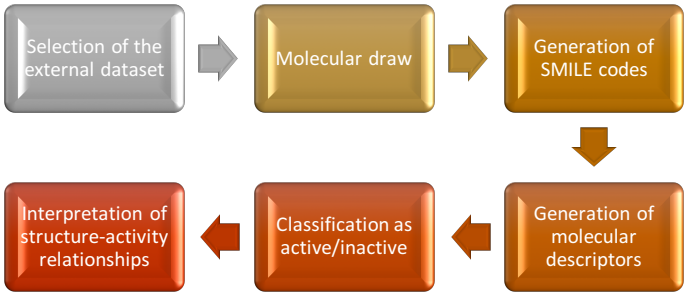
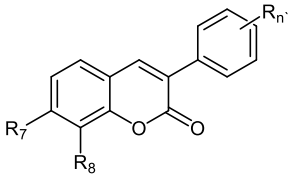


Figure 1. QSAR methodology.

An external dataset was prepared (**Table 1**), including a total of 15 synthetic coumarins prepared by the research group.

Table 1. External dataset.

Compound							SMILE code
1	**R_7**	**R_8**	**$R_{2'}$**	**$R_{3'}$**	**$R_{4'}$**	**$R_{5'}$**	OC1=C(OC(C(C2=CC=CC=C2)=C3)=O)C3=CC=C1
2	H	OH	H	H	H	H	OC1=C(OC(C(C2=CC=C(O)C=C2)=C3)=O)C3=CC=C1
3	H	OH	H	H	OH	H	OC1=C(OC(C(C2=CC=CC(O)=C2)=C3)=O)C3=CC=C1
4	H	OH	H	OH	H	H	OC1=C(OC(C(C2=CC=C(O)C(O)=C2)=C3)=O)C3=CC=C1
5	H	OH	H	OH	OH	H	OC1=C(OC(C(C2=CC=CC=C2O)=C3)=O)C3=CC=C1
6	H	OH	OH	H	H	H	OC1=C(OC(C(C2=CC=C(C)C=C2)=C3)=O)C3=CC=C1
7	H	OH	H	H	Me	H	O=C1OC2=CC(O)=CC=C2C=C1C3=CC=C(C)C=C3
8	OH	H	H	H	Me	H	O=C1OC2=C(O)C(O)=CC=C2C=C1C3=CC=CC=C3
9	OH	OH	H	H	H	H	O=C1OC2=CC(O)=CC=C2C=C1C3=CC=CC(O)=C3
10	OH	H	H	OH	H	H	O=C1OC2=C(O)C(O)=CC=C2C=C1C3=CC(O)=CC=C3
11	OH	OH	H	H	H	OH	O=C1OC2=C(O)C(O)=CC=C2C=C1C3=CC=CC=C3O
12	OH	OH	OH	H	H	H	O=C1OC2=C(O)C(O)=CC=C2C=C1C3=CC(O)=C(O)C=C3
13	OH	OH	H	H	OH	OH	O=C1OC2=CC(O)=CC=C2C=C1C3=CC(Br)=CC=C3
14	OH	H	H	H	H	Br	O=C1OC2=C(O)C(O)=CC=C2C=C1C3=CC=C(O)C=C3
15	OH	OH	H	H	OH	H	O=C1OC2=CC(O)=CC=C2C=C1C3=CC=CC=C3O
16	OH	H	OH	H	H	H	O=C1OC2=CC(O)=CC=C2C=C1C3=CC=CC=C3

Results and discussion. The probability of 3-arylcoumarins (Table 1) being active or inactive was determined (Table 2). The external dataset comprises molecules with hydroxyl substitutions at positions 7 and 8 of the coumarin ring. Other substituents as methyl groups or halogen atoms, present in the 3-phenyl ring, were also considered. Several research groups have explored the importance of the number and position of the different substituents in the coumarin nucleus, considering steric volume, lipophilicity and/or electronic properties, to improve activity and selectivity. However, few studies have been done regarding the relationship between the coumarin structure and clastogenicity.

The results obtained in the present study show, that for this series of compounds, the *in silico* genotoxicity seems to be related to the polyhydroxylated compounds, or where the hydroxyl substituent is at 7 and 8 position, and there is also a catechol on the 3-phenyl ring (compound **12**). It should be noted that the external data set is small and limits an exhaustive SAR. It seems that compounds substituted by hydroxyl groups at position C2' could be clastogenic (compound **11**), if the C7:C8 hydroxyl groups are present. If only one hydroxyl is present in one of these two positions, the compounds are classified as inactive, i.e. comparing of compounds **3** and **9**, and **5** and **15** (Tables 1 and 2).

Substitutions at position 7 (along with 3) of the coumarin ring have been the most studied against various pharmacological activities. Previous SAR studies on coumarin derivatives have suggested that selectivity is determined primarily by the nature of the substituent at position 7 of the coumarin ring. Despite these

evidences, this study shows the importance for the clastogenicity of substitutions at position 7, so it is suggested to extend this study *in silico* and validate it experimentally.

Table 2. *In silico* prediction of clastogenic activity.

Compound	Probability (%)	Classification *
1	75.7	G1:-1
2	67.1	G1:-1
3	67.0	G1:-1
4	52.7	G1:-1
5	64.4	G1:-1
6	79.4	G1:-1
7	82.1	G1:-1
8	62.6	G1:-1
9	70.5	G1:-1
10	52.4	G1:-1
11	50.6	G2:1
12	62.7	G2:1
13	74.1	G1:-1
14	52.2	G1:-1
15	68.1	G1:-1
16	70.6	G1:-1

* G1:-1 = inactive; G2:1 = active.

In the present study, the methyl groups appear to contribute greatly to the increase of the probability of being inactive, independent of the position where the hydroxyl groups are (compounds **6** and **7**), likewise the presence of a halogen in the 3-phenyl ring contributes to inactivity.

In a general, it is necessary to incorporate in this series other compounds of design that allow a more comprehensive SAR interpretation. However, a tendency to clastogenicity of polyhydroxyl-3-arylcoumarins is observed (Figure 2).

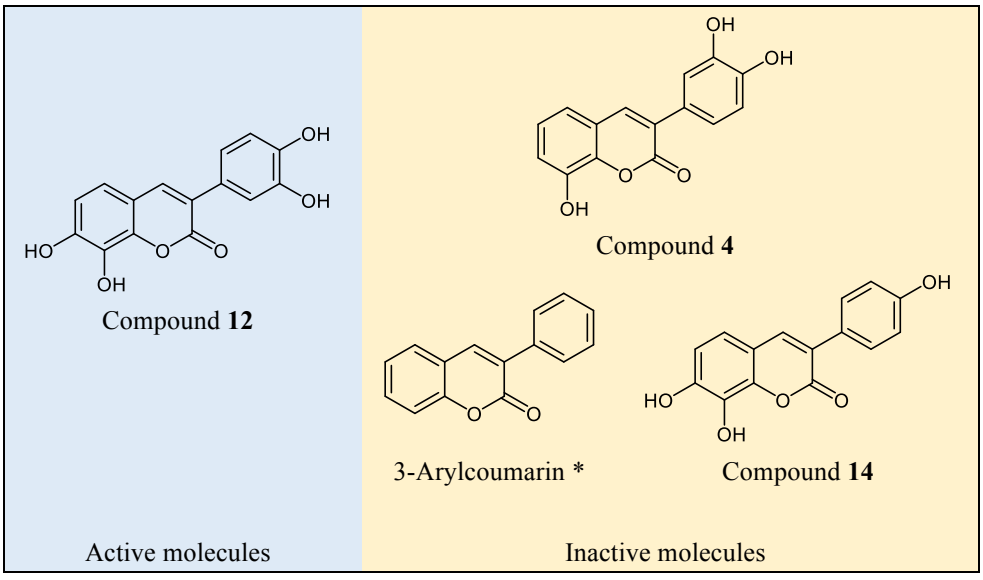
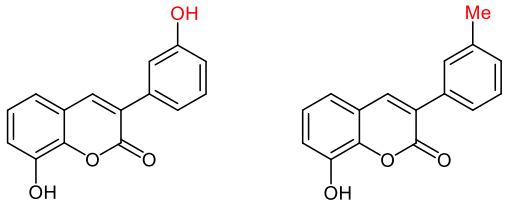


Figure 2. Example of *in silico* 3-arylcoumarins predicted as active/inactive. * prediction of this coumarin was previously published [14,15].

This *in silico* results, specially for compounds **2** and **6** that presented high antioxidant activity in ORAC-FL studies, are significant for the future work of the research group [11]. These compounds could become future antioxidants for several uses, if these theoretical results are validated as they have been classified as non-clastogenic. From the structural point of view, these 8-hydroxy-3-arylcoumarins have differences, given by the presence of the methyl group instead of hydroxyl at C3' position (Figure 3). There is experimental evidence to suggest that the presence of methyl groups in coumarins makes the molecule less toxic. As an example, Maistro et al. (2011) studied that the antigenotoxic and anticlastogenic activities of 4-methylcellulose (synthetic coumarin) could be a result of its antioxidant activity. These authors suggested that the reduction of free radical formation may prevent genetic damage [17]. This structural fragment must be considered in the design of new derivatives, since the clastogenic model used in previous studies carried out with natural coumarins recognizes the methyl substituents as an important chemical feature for the compounds to be non-clastogenic [14,15,18,19].



Compound **2** Compound **6**

Figure 3. 3-Arylcoumarins with high *in vitro* antioxidant activity (predicted as non clastogenic *in silico*).

Conclusions. The preliminary interpretation of the relationship between structure and clastogenicity, considering this external dataset, suggests the importance of hydroxyl groups in the coumarin ring at 7 and/or 8 positions. In general, a tendency to be clastogenic was observed for polyhydroxylated coumarins (specially positions 7 and 8), together with a catechol group in the 3-phenyl group. Future work of SAR is required, for which it is necessary to expand the dataset and the experimental studies that allow to validate these predictions.

References

1. De Surya, K.; Richard, A. *Synthesis* **2005**, *8*, 1231.
2. Hoult, J.R.S.; Payá, M.G. *Pharmacol* **1996**, *27*, 713.
3. Santana, L.; Uriarte, E.; Roleira, F.; Milhazes, N.; Borges, F. *Curr Med Chem* **2004**, *11*, 3239.
4. Accame, M. *Panorama Actual Med* **2000**, *24*, 432.
5. Matos, M.J.; Vazquez-Rodriguez, S.; Fonseca, A.; Uriarte, E.; Santana, L.; Borges, F. *Curr Org Chem* **2017**, *21*, 311.
6. Muñoz, A.; Fonseca, A.; Matos, M.J.; Uriarte, E.; Santana, L.; Borges, F.; Figueroa , R.; Olea-Azar, C. *ChemistrySelect* **2016**, *1*, 4957.
7. Matos, M.J.; Mura, F.; Vazquez-Rodriguez, S.; Borges, F.; Santana, L.; Uriarte, E.; Olea-Azar, C. *Molecules* **2015**, *20*, 18264.
8. Figueroa-Guinez, R.; Matos, M.J.; Vazquez-Rodriguez, S.; Santana, L.; Uriarte, E.; Olea-Azar C; Borges, F.; Diego Maya, J. *Curr Top Med Chem* **2015**, *15*, 850.
9. Figueroa Guíñez, R.; Matos, M.J.; Vazquez-Rodriguez, S.; Santana, L.; Uriarte, E.; Olea-Azar , C.; Maya, J.D. *Fut Med Chem* **2013**, *5*, 1911.

10. Vazquez-Rodriguez, S.; Figueroa-Guíñez, R.; Matos, M.J.; Santana, L.; Uriarte, E.; Lapier, M.; Maya, J.D.; Olea-Azar, C. *MedChemComm* **2013**, *4*, 993.
11. Matos, M.J.; Pérez-Cruz, F.; Vazquez-Rodriguez, S.; Uriarte, E.; Santana, L.; Borges, F.; Olea-Azar, C. *Bioorg Med Chem* **2013**, *21*, 3900.
12. Venugopala, K.N.; Rashmi, V.; Odhav, B. *BioMed Res Inter* **2013**, *2013*, 1.
13. Kaneko, T.; Baba, N.; Matsuo, M. *Cytotechnology* **2001**, *35*, 43.
14. Guardado-Yordi, E.; Matos, M.J.; Uriarte, E.; Santana, L.; Abreu, O.; Molina, E. In *Natural coumarins: Qstr approaches regarding their genotoxicity*, 18th ECSOC, Sciforum Electronic Conference Series, 2014; 18, e015: Sciforum Electronic Conference Series.
15. Guardado-Yordi, E.; Matos, M.J.; Santana, L.; Uriarte, E.; Abreu, O.; Molina, E. *Biol Chem Res* **2015**, 241.
16. Estrada, E.; Molina, E. *J Mol Graph Mod* **2006**, *25* 275.
17. Maistro, E.L.; Angeli, J.P.F.; Andrade, S.F.; Mantovani, M.S. *Gen Mol Res* **2011**, *10*, 1130.
18. Guardado-Yordi, E.; Matos, M.J.; Castro, R.; Santana, L.; Uriarte, E.; Molina, E. *LWT - Food Sci Techn* **2015**, *61*, 216.
19. Guardado-Yordi, E.; Molina, E.; Matos, M.J.; Uriarte, E. *J Biomol Screen* **2012**, *17*, 216.