QSAR & Combinatorial Science QCS

PLS Analysis for Antibacterial Activity of Natural Coumarins Using VolSurf Descriptors

Catherine Koukoulitsa,[a]* Anna Tsantili-Kakoulidou,[b] Thomas Mavromoustakos,[a,c] Ioanna Chinou[d]

[a] Institute of Organic and Pharmaceutical Chemistry, National Hellenic Research Foundation, 48 Vas. Constantinou Avenue, 11635 Athens, Greece
*E-mail: koukoulitsa@eie.gr
[b] Department of Pharmaceutical Chemistry, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, 157 71, Athens, Greece
[c] Department of Chemistry, University of Athens, Panepistimiopolis Zographou 15771 Athens, Greece
[d] Department of Pharmacognosy & Natural Products, School of Pharmacy, University of Athens, Panepistimiopolis, Zografou, 157 71, Athens, Greece

Keywords: Antibacterial activity, Natural coumarins, Multivariate data analysis, VolSurf descriptors, Semiempirical calculations, Natural products

Presented at the EuroQSAR Conference

Received: December 15, 2008; Accepted: February 13, 2009

DOI: 10.1002/qsar.200860189

Abstract
In the present study the antibacterial activity of a dataset including 24 natural coumarin derivatives was investigated by means of multivariate data analysis. Antibacterial data were available against six pathogenic microorganisms. VolSurf descriptors, used to characterize structural properties, proved efficient to discriminate the compounds according to different levels of activity as shown in the PCA and PLS-DA score plots. An informative PLS model considering the six response variables in parallel with satisfactory statistics was obtained, based on 39 original descriptors ($R^2 = 0.86$ and $Q^2 = 0.79$). Descriptors related to the hydrophilic regions and to hydrogen bonding capability proved to exert the most significant influence on projection ($VIP > 1.5$) most of them contributing in a positive way to the antibacterial activity. Contrary, lipophilicity expressed as log$P$ was found to have a negative influence. Elongation of the molecules as described by EEFR is an additional unfavorable structural feature.

1 Introduction

Coumarins represent a category of heterocyclic natural compounds that are widely distributed in the plant kingdom [1]. Very interesting subclasses of this category are the linear or angular pyrano and furanocoumarins that have been found to possess antiproliferative, antiviral and antibacterial activities. They are characterized by a variety of oxygenation patterns on the benzopyrone nucleus and display a remarkable array of biochemical and pharmacological actions. Up to now literature reports concerning quantitative structure-activity relationships for the antibacterial activity of coumarins are very limited. In the last years however QSAR studies are progressively extended to natural products [2–5], in the aim to obtain models, which could be used as tool for activity prediction of newly isolated compounds or to direct synthetic chemistry to exploit better the information incorporated in the compounds provided by nature. Considering the arsenal of molecular parameters which can be used in QSAR studies VolSurf descriptors provide a global depiction of the chemical structure in respect to molecular size and shape, to size and shape of both hydrophilic and hydrophobic regions and to the balance between them. VolSurf is a computational procedure that is actually designed to produce descriptors related to pharmacokinetic properties, starting from 3D molecular field maps. In the standard procedure, GRID interaction fields [6] are calculated around the target molecules. The basic concept of VolSurf is to compress the information present in 3D grid maps into few 2D nu-

Supporting information for this article is available on the WWW under www.qcs.wiley-vch.de

merical descriptors, which are simple to understand and to interpret [7, 8].

In the present study the applicability of the VolSurf descriptors in respect to the antibacterial properties of a dataset of natural coumarins was investigated in the aim to establish QSAR models, which would provide information on the structural features influencing the activity. For this purpose multivariate data analysis was used, applying a statistical methodology, which can handle a large number of interrelated descriptors, exploiting the maximum information encoded within them [9]. The investigated natural coumarins had previously been isolated from the plants *Seseli devenyense* and *Peucedanum luxurian* and *MIC* values were determined against Gram-positive bacteria (*S. aureus*, *S. epidermidis*) and Gram-negative bacteria (*E. coli*, *E. cloacae*, *K. pneumoniae*, *P. aeruginosa*) [10].

2 Methods

2.1 Data Set

The data set comprises 24 natural coumarins. They are linear pyrano or angular furanocoumarins or they possess oxygenated substituents, in some cases including a glycoside moiety, on the simple coumarinic skeleton. Their chemical structures of the investigated compounds are presented in Figure S1 in the Supplementary Material.

Their antibacterial activity data against *S. aureus*, *S. epidermidis*, *E. coli*, *E. cloacae*, *K. pneumoniae*, and *P. aeruginosa* were taken from reference [10] and were expressed as log (1/*MIC*) (Table S1 Supplementary Material). Six compounds were inactive and *MIC* values could not be determined at the conditions used in the assays. In those cases an indicative value, lower than that of the weakest active compound, was assigned (log (1/*MIC*) = 1.2), in order to include them in the analysis.

2.2 Molecular Descriptors Calculation

Molecular structures of compounds were generated using SYBYL molecular modeling package [11], and their energies were minimized using the Powell method with a convergent criterion provided by the Tripos force field [12]. As a next step, the VolSurf software [7, 8] was used to generate 94 descriptors, setting H_2O, DRY, and O probes to characterize structural properties. Lipophilicity (log*P*) was computed by means of a linear equation derived by fitting VolSurf descriptors to experimental data on water/octanol partition coefficient. The meaning of VolSurf descriptors is described in Table S2, Supplementary Material.

2.3 Multivariate Data Analysis

Simca-P 10.5 (Umetrics, California, USA) was used to perform principal component analysis (PCA), partial least square analysis (PLS) and PLS-discriminant analysis (PLS-DA) [9]. The pool of VolSurf descriptors formed the **X** matrix. The six sets of antibacterial data, expressed as log (1/*MIC*), were introduced as response variables in the **Y** matrix. Variable selection was based on the Variable Influence on Projection criterion and on the size of the coefficients of the variables.

3 Results and Discussion

3.1 Data Overview

Principal component analysis was applied to the **X**-matrix of the 94 VolSurf descriptors. A 4 component PCA-X model was obtained with $R^2 = 0.82$ and $Q^2 = 0.69$. The score plot of the first two components showed a tendency to discriminate the compounds according to their activity and their structural characteristics (Figure not shown). As a next step PLS-DA was performed, classifying the compounds according to three levels of activity: active, weakly active, and inactive. As active compounds were considered those with log (1/*MIC*) = 2.10 – 3.92, as weakly active those with log (1/*MIC*) = 1.41 – 2.01. One compound from each level (**2**, **16**, **11**), as well as some border cases (**13**, **14**, **23**), were considered as test set. The 3 component PLS-DA model with $R^2 = 0.85$ and $Q^2 = 0.76$ provided a clear discrimination with the active compounds scattered in the two right quartiles and the inactive located in the upper left quartile. The weakly active compounds form two clusters in the left low quartile, one clearly separated and a second closer to the active compounds. The test compounds **2**, **16**, **11** are correctly classified. Compounds **13** and **14** were found relatively close to the active analogs, while compound **23** although with a relatively high log (1/*MIC*) value, belonged to the second cluster of the weak active compounds (Fig. 1).

The six response variables were found to be strongly interrelated (PCA-Y model: $A = 2$, $R^2 = 0.99$, $Q^2 = 0.97$) and were considered in parallel for PLS analysis. According to the loading plot, however, activity data formed two separate clusters. Cluster I includes *P. aeruginosa*, *K. pneumoniae* and *E. coli*, and cluster II includes *S. aureus*, *S. epidermis* and *E. cloacae*.

3.2 PLS Analysis

PLS analysis was performed using the same training and test sets as described for PLS-DA. A two component PLS model based on the 94 VolSurf descriptors was obtained with overall $R^2 = 0.83$ and $Q^2 = 0.70$. After variable selection according to the variable importance to projection criterion (*VIP*) and the magnitude of the coefficients a refined PLS model based on 39 original descriptors was obtained with overall $R^2 = 0.86$ and $Q^2 = 0.79$. The model was validated for its robustness applying the permutation test

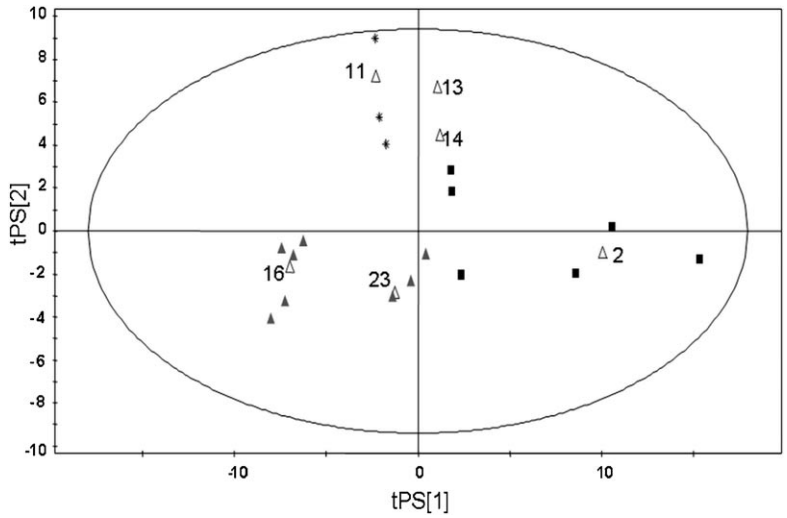

Figure 1. PLS-DA score plot. ■ active, ▲ weak active, * inactive, △ test compounds.

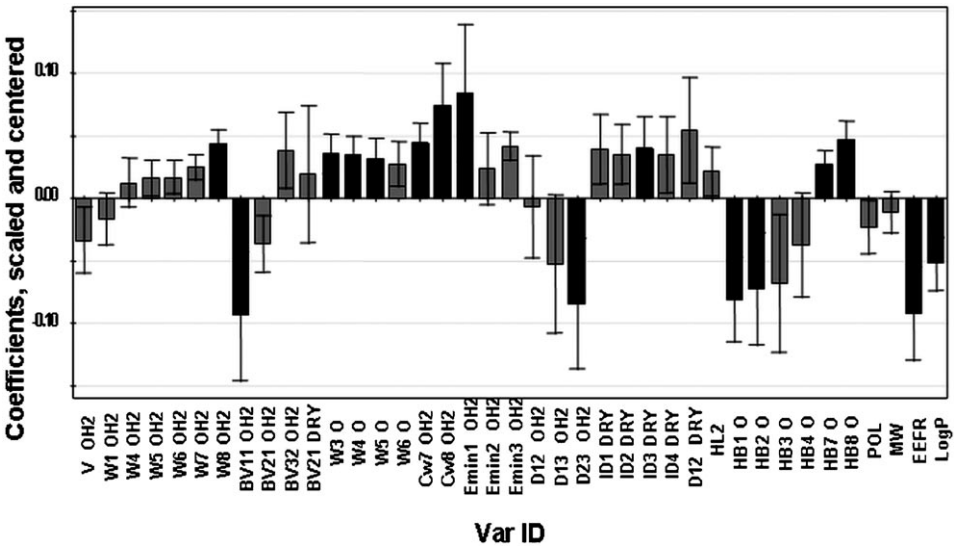

Figure 2. Coefficients histogram as generated for antibacterial data against *S. aureus*. In black are depicted descriptors with $VIP > 1$.

(20 runs) implemented in SIMCA, as well as re-analyzing the scrambled **Y** data using the whole pool of **X**-matrix descriptors [13].

As shown in Figure 2 descriptors related to the hydrophilic regions and to hydrogen bonding capability proved to exert the most significant influence on projection ($VIP > 1.5$), most of them contributing in a positive way to the antibacterial activity. Contrary, lipophilicity expressed as log*P* was found to have a negative influence. Elongation of the molecules as described by EEFR is an additional unfavorable structural feature for antibacterial activity of the investigated coumarin derivatives. The hydrophilicity requirements for the antibacterial activity of the coumarin derivatives, in particular in the substitution pattern attached to the coumarin skeleton, can be visualized in Figure 3. Figure 3 illustrates the hydrophilic regions contour maps at -0.2 kcal/mol of a) the most active compound **6**, and b) the inactive compound **12**. The extended elongation of the inactive compound **12** is characteristically depicted.

Considering the model statistics for the individual response variables, *S. aur*eus, *S. epidermis*, and *E.cloacae* belonging to cluster II, show better statistical data than *P. aeruginosa, K. pneumoniae*, and *E. coli*, which form cluster I. Separate analysis of Cluster II response variables resulted in a further refined 2 component model with $R^2 = 0.89$ and $Q^2 = 0.85$. In Figure 4a, 4b the plot of observed versus predicted -log MIC data is presented for *E.cloacae* and *S.*

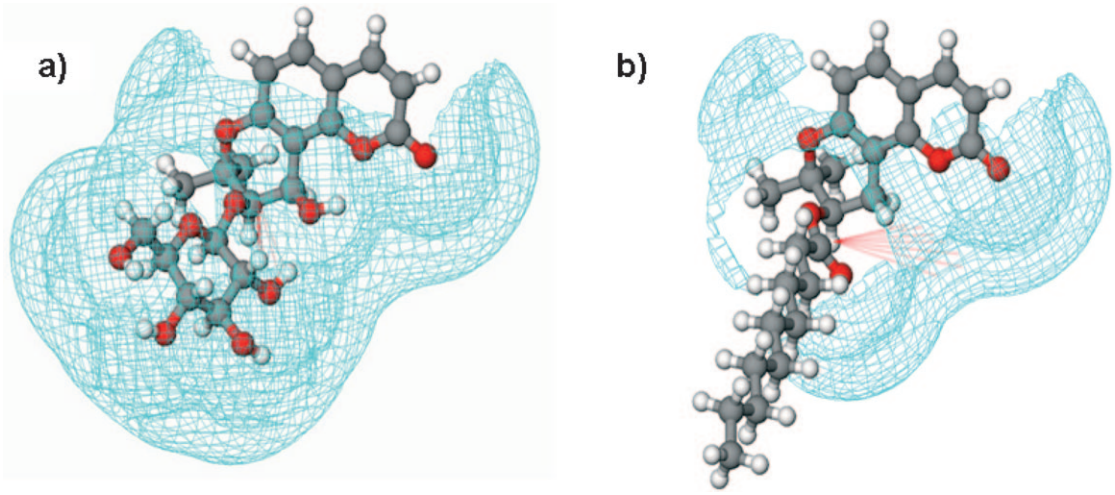

Figure 3. a) Hydrophilic regions contoured at −0.2 kcal/mol of the most active compound **6**, and **b)** the inactive compound **12.**

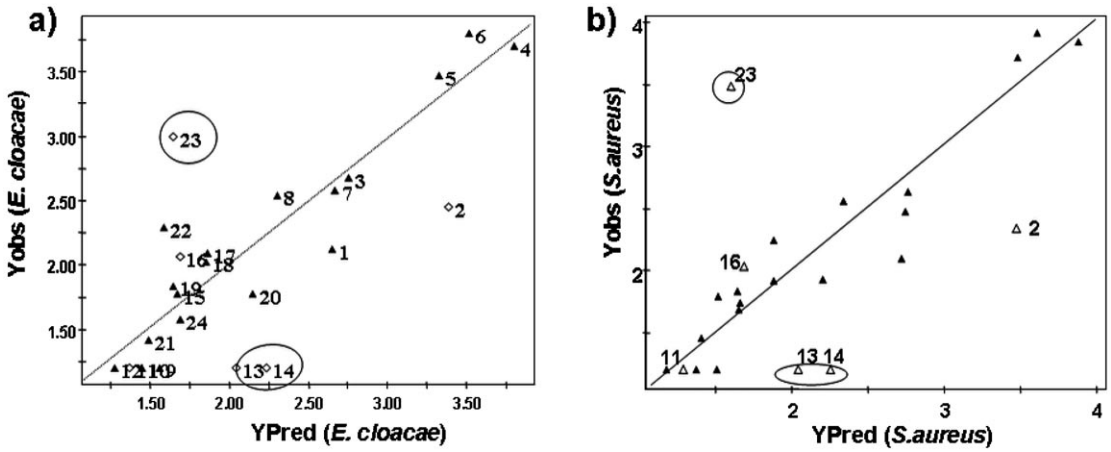

Figure 4. Observed versus predicted log (1/*MIC*) data for **a)** *E. cloacae* and **b)** *S. aureus.*

aureus, respectively. For compound **23** weak activity was predicted, while for the inactive compounds **13** and **14** weak activity was assigned. Taking into account the structural features of these compounds, reflected also in the PCA and PLS- DA score plots, a reconsideration of their biological activity data is suggested.

4 Conclusions

VolSurf descriptors were successfully applied in discriminating the antibacterial activity of the natural coumarins in different levels as a result of structural characteristics and were efficient to produce a QSAR model for further predictions. Hydrophilic regions proved to be most important for activity. These results suggest that the applicability domain of VolSurf descriptors could be extended to pharmacodynamic processes.

5 Acknowledgements

Catherine Koukoulitsa completed this research activity while she was a recipient of IKY (State Scholarships Foundation). The authors are grateful to Prof. Gabriele Cruciani (Laboratory for Chemometrics, School of Chemistry, University of Perugia, Italy) for kindly providing us the VolSurf program.

6 References

[1] R. D. H. Murray, in *Progress in the Chemistry of Organic Natural Products* (Eds: W. Hertz, H. Falk, G. W. Kirby, R. E. Moore), Springer, Berlin **2002**, pp. 1 – 529.
[2] G. Ramírez-Galicia, R. Garduño-Juarez, B Hemmateenejad, O. Deeb, S. Estrada-Soto, *Chem. Biol. Drug Des.* **2007**, *70*, 143 – 153.

[3] A. D. Wright, R. de Nys, C. K. Angerhofer, J. M. Pezzuto, M. Gurrath, *J. Nat. Prod.* **2006**, *69*, 1180 – 1187.
[4] I. R. A. Menezes, J. C. D. Lopes, C. A. Montanari, G. Oliva, F. Pavão, M. S. Castilho, P. C. Vieira, M. T. Pupo, *J. Comput. Aid. Mol. Des.*, **2003**, *17*, 277 – 290.
[5] Zaheer-ul-Haq, B. Wellenzohn, S. Tonmunphean, A. Khalid, M. Iqbal Choudhary, B. M. Rodea, *Bioorg. Med. Chem. Lett*., **2003**, *13*, 4375 – 4380.
[6] P. J. Goodford, *J. Med. Chem*., **1985**, *28*, 849 – 857.
[7] G. Cruciani, M. Pastor, W. Guba, *Eur. J. Pharm. Sci.*, **2000**, *11*, *Suppl. 2*, S29 – S39.
[8] G. Cruciani, P. Crivori, P.-A. Carrupt, B. Testa, *J. Mol. Str.: THEOCHEM*, **2000**, *503*, 17 – 30.
[9] L. Eriksson, E. Johansson, N. Kettaneh-Wold, S. Wold, *Multi-and Megavariate Data Analysis, Principles and Applications*, Umetrics AB, Umea, Sweden **2001**.
[10] J. Widelski, K. Craikou, K. Glowniak, E. Chinou, I. Chinou, *Int. J. Antimicrob.* Agents, **2007**, *29*, *Suppl. 2*, S475.
[11] J. G. Vinter, A. Davis, M. R. Saunders, *J. Comput. Aided Mol. Des.*, **1987**, *1*, 31 – 51.
[12] *SYBYL*, version 8.0, TRIPOS Associates Inc. S.Louis MO.
[13] C. Rücker, G. Rücker, M. Meringer, *J. Chem Inf. Model.* **2007**, *47*, 2345 – 2357.

