QSAR Study of Antimicrobial Activity of Some 3-Nitrocoumarins and Related Compounds†

Željko Debeljak,‡ Armin Škrbo,§ Ivona Jasprica,|| Ana Mornar,|| Vanda Plečko,⊥ Mihajlo Banjanac,# and Marica Medić-Šarić*,||

Department of Medicinal Biochemistry, Clinical Hospital Osijek, J. Huttlera 4, 31000 Osijek, Croatia, Department of Medical Informatics, Faculty of Pharmacy, University of Sarajevo, Čekaluša 90, 71000 Sarajevo, Bosnia and Hercegovina, Department of Medicinal Chemistry, Faculty of Pharmacy and Biochemistry, University of Zagreb, A. Kovačića 1, 10000 Zagreb, Croatia, Department of Clinical and Molecular Microbiology, Clinical Hospital Center Zagreb, Šalata 2., 10000 Zagreb, Croatia, and GlaxoSmithKline Research Centre Zagreb, Prilaz baruna Filipovića 29, 10000 Zagreb, Croatia

Received October 31, 2006

A new class of antimicrobial agents, 3-nitrocoumarins and related compounds, has been chosen as a subject of the present study. In order to explore their activity and molecular properties that determine their antimicrobial effects, QSAR models have been proposed. Most of the 64 descriptors used for the development were extracted from semiempirical and density functional theory (DFT) founded calculations. For this study literature data containing results of microbiological activity screening of 33 coumarin derivatives against selected clinical isolates of *C. albicans* (CA) and *S. aureus* (SA) have been selected. Multivariate predictive models based on random forests (RF) and two hybrid classification approaches, genetic algorithms (GA) associated with either support vector machines (SVM) or k nearest neighbor (kNN), have been used for establishment of QSARs. An applied feature selection approach enabled two-dimensional linear separation of active and inactive compounds, which was a necessary tool for rational candidate design and descriptor relevance interpretation. Candidate molecules were checked by cross-validated models, and selected derivatives have been synthesized. Their antimicrobial activities were compared to antimicrobial activities of the representative derivatives from the original set in terms of minimal inhibitory concentration (MIC) against chosen SA and CA ATCC strains. High ranking of descriptors consistent with the degree of hydrolytic instability of selected compounds is common to models of antimicrobial activity against both microorganisms. However, descriptor ranking indicates different antimicrobial mechanisms of action of chosen coumarin derivatives against selected microbial species.

1. INTRODUCTION

Scientific interest for antimicrobial activity of coumarins continuously grows from the mid 20th century until today.[1−11] As a result, a few coumarin antibiotics became candidates for human and veterinary medicine applications. The most important representative is the 3-aminocoumarin derivative novobiocin, the antibiotic that has been relatively recently approved for medical use in the U.S. for SA infection treatment.[12] Besides novobiocin, other coumarin derivatives like eskuletin, umbelliferon, and related compounds possess antibacterial properties as well.[13] Antifungal activity has been attributed to some of the coumarin derivatives, including coumarin (1,2-benzopyranone) itself.[7,14]

Antimicrobial activity of 3-nitrocoumarins and related compounds against CA and SA has been shown relatively recently.[6,15,16] Although it has been confirmed that 3-nitrocoumarin inhibits growth of CA, only indirect evidence exists that the growth inhibition is accomplished by irreversible phosphatidylinositol phospholipase C (PI-PLC) inhibition.[16] On the other hand, these compounds share the coumarin backbone with novobiocin, which inhibits DNA gyrase. However, this enzyme is not a probable molecular target for antimicrobial action of these compounds since they lack a 7-OH or similar substituent at that position.[2,17] Such a substituent interferes with the attachment of an ATP molecule to the DNA gyrase, and it also enables formation of the H-bond between the DNA gyrase and coumarin molecule. Still, moderate activity of these compounds especially against CA represents a good starting point for the modeling of antimicrobial activity. By changing substituents at positions 2 and 3, and most importantly at position 4, significant modulation of antimicrobial activity against both microorganisms has been achieved.[15] These results seemed suitable for the following QSAR study. It has been estimated that there is room for further improvement of activity of these compounds against SA and CA. Machine and statistical learning approaches have been selected for the task of new candidate design. Besides inherent feature selection capabilities, selected approaches provided the tool for antimicrobial activity prediction.

† Dedicated to Professor Nenad Trinajstić on the occasion of his 70th birthday.

* Corresponding author phone: +385 (0) 1 4818 304; fax: +385 (0) 1 4856 201; e-mail: bebamms@pharma.hr.

‡ Clinical Hospital Osijek.

§ University of Sarajevo.

|| University of Zagreb.

⊥ Clinical Hospital Center Zagreb.

GlaxoSmithKline Research Centre Zagreb.

10.1021/ci600473z CCC: $37.00
Published on Web 05/10/2007

2. MATERIALS AND METHODS

2.1. Data Set. The selected set composed of 33 3-nitrocoumarins derivatives and related compounds classified as active (+) or inactive (−) on the basis of corresponding microbiological screening[6,15] is given in Table 1. All of the selected compounds are substituted at positions 2, 3, and 4. The majority of substituents at position 4 are heterocycles that vary in size and composition. Besides 3-nitro derivatives, some 3-amino derivatives are also included. Compounds 29−32 are tautomers of compounds 10−13.[6] Designed structures along with predicted activity classes are given in the last part of Table 1.

Microbiological activity of compounds from Table 1 that were screened by the well-diffusion method has been taken from the literature.[6,15] Authors[6,15] encountered problems related to the low solubility and hydrolysis of some molecules from the given set which disabled the adequate determination of corresponding minimal inhibitory concentrations (MIC). Well-diffusion based antimicrobial activity screening is less affected by these problems. Therefore Škrbo[15] and Govori et al.[6] have chosen this approach. However, diameters of inhibition zone (DIZ) that they published represent only a rough estimation of antimicrobial activity. In order to lessen the problems related to the bias or variability inherent to DIZ measurements taken from literature[6,15] molecules from Table 1 were divided in two activity classes. Substances with measurable DIZ (DIZ $\geq$7 mm) at given concentrations have been assigned to an active class ("+"). Otherwise, substances were regarded as inactive ("−").

Preliminary results have shown that 3-nitrocoumarins possess considerable antifungal activity. Therefore antimicrobial screening results of nine clinical isolates of CA were used for the modeling of activity against this microorganism. For the purpose of a model development results corresponding to all nine strains were aggregated. If a substance proved to be inactive against any of the analyzed strains, it was assigned to the inactive class. Otherwise it was assigned to the active class. Clinical isolates of SA proved to be more resistant. In order to increase variability of results needed for appropriate model development in the case of SA a higher dose was used.

2.2. Descriptors Generation. Sixty-four attributes have been generated for the description of selected coumarin derivatives. Quantum chemical descriptors like Mulliken charges (MC), total valences (TV), bond orders (BO), and lengths (R) which pertain to the coumarin ring and attached substituents enable pharmacophore assembly after the feature selection process has occurred. Calculation of quantum chemical descriptors was preceded by molecular geometry optimization based on the PM3 semiempirical approach. After detailed conformation analysis structures with the lowest total energy were used for single point B3LYP 6-31G** based electronic properties calculations. Both semiempirical and DFT calculations were carried out by GAMESS-US[18] release 22. 11. 2004. for in vacuo systems.

More computationally demanding quantum chemical approaches to the descriptor generation were considered during the preliminary phase of this study. Inclusions of the solvent effect based on the polarizable continuum model or periodic boundary box approach were examined. Unfortunately, the central processing unit time needed for these calculations proved to be too long. Moreover, serious problems related to the final geometry arose. Namely, steric hindrance between the S3 and S4 and some intramolecular H-bonds were detected in some cases. Due to the steric hindrance, the possibility of finding the global energy minimum is determined by the selection of initial molecular geometry. This problem could not be solved simply by application of a different optimization algorithm. Therefore the influence of the corresponding torsion angles posed the need for even more complex calculations. This demand additionally increased computational burden. Only application of the semiempirical in vacuo approach made all these calculations feasible for more than 40 molecules presented in this study. This approach included iterative torsion angle variation which resolved the problem of steric hindrance which obstructed geometry optimization. For DFT calculations the GAMESS package provides additional quantities that are not present among PM3 calculation results. In order to enlarge the descriptor pool, single point DFT calculations, which are not very demanding in terms of processing time, were made for molecules for which geometry was optimized by the PM3 approach.

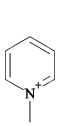
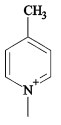
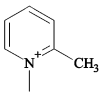
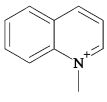
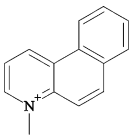
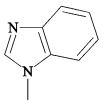
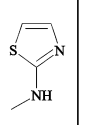
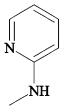
Besides quantum chemical descriptors, five more descriptors were included in the pool: ClogP,[19] molar refractivity (MR),[20] molecular mass (Mr), solvent accessible volume, and area.[21] Due to the space requirements the complete descriptor list and matrix containing corresponding values are omitted. However, they are available on request.

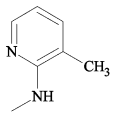
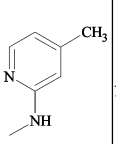
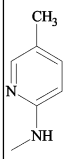
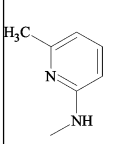
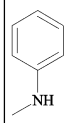
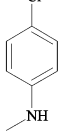
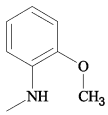
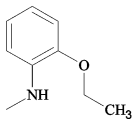
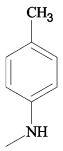
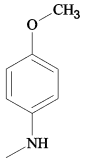
2.3. Learning Tools. Three machine and statistical learning tools were utilized in this study for QSAR model development: RF,[22,23] GA/SVM,[24] and GA/kNN.[25] Since all techniques have been thoroughly described in the provided references, a detailed method description has been omitted. However, implementation details specific for this study are given in the following text.

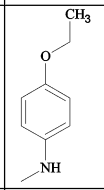
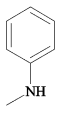
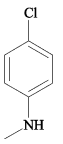
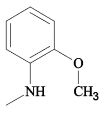
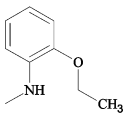
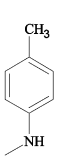
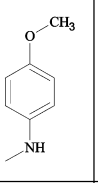
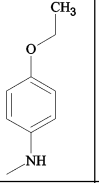
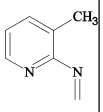
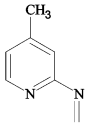
Svetnik and coauthors[23] proposed near optimal values of the two most important RF user-defined variables: number of trees and number of descriptors used for growth of each tree. For classification tasks they assigned a value of 500 to the first variable and square root of the total descriptor number as the first choice for the second variable. We used these predefined values during the development of RF based QSAR models presented in this study.

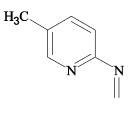
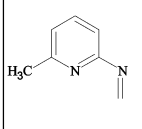
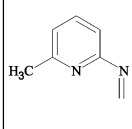
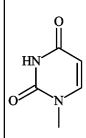
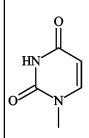
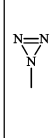
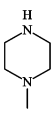
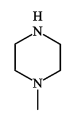
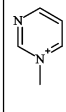
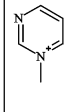
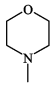
GA/SVM represents a more demanding learning approach. Our implementation presumes predefinition or selection of 7 user-defined variables. The selection is realized by an internal leave-one-out (LOO) loop nested within an external validation LOO loop.[26,27] This procedure is described in the following text. Since we applied a radial basis function SVM kernel, two SVM parameters had to be adjusted (C, γ). In order to save resources the internal loop was used for fine-tuning of only these two SVM specific variables. $\log_2$C was searched across the [−5, 15] interval, while $\log_2\gamma$ was searched across [−15, 5]. An iteration step equal to 5 was used for the tuning of both variables. Point mutation distribution skewness was set to 10 in the GA/SVM case.[28,29] All other GA specific variables were predefined and kept fixed throughout the learning procedure (Table 2.). After the training, only the best performing chromosome from the last generation was used for external validation and prediction purposes.

Table 1. 3-Nitrocoumarins and Related Compounds[f]

	Molecule No.	1	2	3	4	5	6	7	8	9	10
SUBSTITUENTS[b]	S4	-Cl									
	S3	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$
	S2	=O	=O	=O	=O	=O	=O	=O	=O	=O	=O
ACTIVITIES	CA[c]	-	+	+	+	+	-	+	+	-	-
	SA	-	-	+	-	-	-	+	+	+	-

	Molecule No	11	12	13	14	15	16	17	18	19	20
SUBSTITUENTS[b]	S4										
	S3	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$
	S2	=O	=O	=O	=O	=O	=O	=O	=O	=O	=O
ACTIVITIES	CA[c]	+	-	-	+	+	+	+	+	+	+
	SA	+	+	+	+	-	-	-	-	-	-

	Molecule No.	21	22	23	24	25	26	27	28	29 (11)[d]	30 (12)
SUBSTITUENTS[b]	S4										
	S3	$-NO_2$	$-NH_2$	$-NH_2$	$-NH_2$	$-NH_2$	$-NH_2$	$-NH_2$	$-NH_2$	$-NO_2$	$-NO_2$
	S2	=O	=O	=O	=O	=O	=O	=O	=O	-OH	-OH
ACTIVITIES	CA[c]	+	+	+	+	+	+	+	+	+	-
	SA	-	+	+	+	+	+	-	+	+	+

	Molecule No.	31 (13)	32 (14)	33	C11[e]	C12	C21	C31	C32	C41	C42	C51	C52
SUBSTITUENTS[b]	S4												
	S3	$-NO_2$	$-NO_2$	$-NO_2$	$-NO_2$	$-NH_2$	$-NO_2$	$-NO_2$	$-NH_2$	$-NO_2$	$-NH_2$	$-NO_2$	$-NH_2$
	S2	-OH	-OH	-OCH3	=O	=O	=O	=O	=O	=O	=O	=O	=O
ACTIVITIES	CA[c]	-	+	+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+	+/+
	SA	+	+	+	+/+	+/+	+/+	+/+	+/+	-/-	+/-	+/+	+/+

[a] Chromen backbone. [b] Substituent label is composed of symbol "S" and the number that assigns the substituent position. [c] If compound lacks activity against at least one of the 9 tested CA strains, then it is proclaimed inactive ("−"). [d] Numbers in brackets mark appropriate tautomeric pairs. [e] Symbol "C" stands for "candidate". These molecules have been designed based on QSAR results. Corresponding predictions of antimicrobial activity obtained by two applied learning tools are given. Tools used for prediction of antimicrobial activity against each microorganism are described in the text. [f] Selected structures and antimicrobial screening results based on the existence or absence of measurable growth inhibition compounds are sorted in two activity classes. Activities against CA and SA correspond to 50 and 100 μL doses, respectively.

Table 2. User-Defined Variables for GA Based Training

variable	value
number of generations	50
number of chromosomes	50
elitism %	20
mutation %	5

GA/kNN was introduced to solve the rare object classification problem. There are only 8 instances when analyzed compounds proved to be inactive against CA (Table 1). Selection of only the first neighbor provides a greater chance for accurate prediction activity of minority objects. Therefore we implemented a GA/1NN ensemble based approach. Ensembles were formed from all chromosomes from the last generation and majority voting was applied. All GA options were kept at the values given in Table 2. Point mutation distribution skewness was set to 1 in this instance.

For prediction of antimicrobial activity against CA, RF and GA/1NN methods were applied, while in the SA case, RF and GA/SVM were used. The use of two or more learning approaches ensures better predictions as well as a more reliable determinants selection.[30] All routines were written in R 2.2.0. programming language,[31] and they are available on request. Calculations were done on a personal computer.

2.4. Feature Selection and Model Validation. Feature selection learning tools provide a means for a rational approach to molecular design and action mechanism analysis. Among the three learning tools applied in this study only RF implements feature selection in a direct fashion.[23] After the training phase was complete, all descriptors were sorted according to Gini's measure. Descriptor rankings summed up across all LOO iterations, and all repetitions of the complete validation process were used for the final selection of biological determinants based on the RF approach.

Both GA/SVM and GA/1NN approaches involved hybridization of GA with the corresponding prediction tool. In this case biological determinants were selected based on the frequency of analyzed descriptor appearance in selected chromosomes. These frequencies were also calculated across LOO iterations and all repetitions of the complete validation process. The only difference between GA/SVM based and GA/1NN based descriptor selection is that the former process involves only the best performing chromosome, while the latter process involves the complete ensemble of chromosomes, i.e., the complete last generation.

Complete validation consists of external LOO validation within which learning is embedded. In the case of GA/SVM and GA/1NN, besides external, there is also an internal LOO loop that is needed for the best chromosomes selection. However, these loops were strictly separated, and the external LOO validation realized this way represents a reliable model evaluation tool.[26,27] There are other external validation protocols available like leave-many-out cross-validation, but a relatively small number of molecules dictates application of LOO.[32]

Applied learning methods involve the usage of random number generators as a part of the learning process. In order to analyze the influence of inherent randomness on the prediction stability, ten repetitions of the complete validation process with different random seeds were made in all cases. Accuracy has been selected for evaluation of predictive performance of a single validation process, while a coefficient of variation (CV) of accuracies obtained across ten repetitions was established as a measure of learning stability. Obtained results were analyzed by statistical tests implemented in Statistica 6.0 (StatSoft, Inc., Tulsa, U.S.A.).

2.5. Laboratory Methods. In order to compare activities of the most potent derivatives from the original set with the most promising candidates and known antibiotics, molecules 8, 23, C51, and C52 were synthesized according to refs 15 and 33. ^{13}C and ^{1}H NMR and FTIR spectra of these compounds are available on request. Corresponding MICs were determined against *C. albicans* ATCC 90028 and *S. aureus* ATCC 29213 strains according to appropriate CLSI guidelines.

3. RESULTS AND DISCUSSION

3.1. SAR. Antimicrobial activities (Table 1) reveal that 3-nitro-4-(imidazol-1-yl)coumarin (compound 7), 3-nitro-4-(benzimidazol-1-yl)coumarin (compound 8), 3-amino-4-anilinocoumarin derivatives (compounds 22−28), and 3-nitro-4-(pyridin-2-ylimino)chromen-2-ol derivatives (compounds 29−33) represent broadly potent molecules. The presence of a quaternary ammonium group (compounds 2−6) does not seem to make either a negative or a positive contribution to antimicrobial activity against any of the analyzed microorganisms. Comparison of substances 2−8 with other derivatives also points out that the existence of an amino linker at position C4 of the coumarin structure is not essential. For example, direct attachment of the 4-imidazol-1-yl substituent exhibits similar or even more pronounced activity. The same group of derivatives implies the importance of S4 size, especially with respect to SA. Bulky substituents like benzoquinolinium significantly degrade activity against analyzed species.

Another interesting substitution is realized by replacement of the 3-nitro group with a 3-amino group in the case of the 4-anilinocoumarins. This replacement does not cause the loss of antimicrobial activity against any of the analyzed microorganisms. In fact, in the case of SA it even enhances antimicrobial activities. Surprisingly, 3-nitro group is not of essential importance for antimicrobial activity against any of the analyzed microorganisms. Still, it seems that substitution at the position 3 modulates the influence of substitution at the position C4 on antimicrobial activity against SA.

Antimicrobial activities of compounds 29−33 imply that the presence of the carbonyl group is also not essential. This fact eliminates all antimicrobial mechanisms that involve a carbonyl group with an H-bond acceptor role. Specifically this refers to DNA gyrase inhibition.[17] Imino derivatives (compounds 29−33) also exhibit significant deviation of the coumarin structure from planarity (due to the space requirements PM3 optimized in vacuo geometries are not shown). Therefore planarity of the coumarin backbone does not play a role in the mechanism of antimicrobial activity. π-stacking which is a part of some coumarin related antimicrobial mechanisms of action is therefore less probable. These conclusions are valid for both microbial species.

In summary, specific structure and, to some extent, size of a substituent at the C4 atom make an appreciable impact on the activity against CA. For example, the presence of the 4-methylpyridin-2-ylamino or the 4-methylpyridin-2-ylimino substituent at C4 diminishes the activity of described

Table 3. Predictive Performance[a]

	microorganism			
	CA		SA	
method	GA/1NN	RF	GA/SVM	RF
<accuracy>%	77.88	79.09	79.39	79.39
CV%	4.87	3.35	5.34	1.61

[a] Presented results are calculated based on ten repetitions of LOO external validation with altered random seed.

derivatives against this microorganism in comparison to their anilino counterparts or 6-methylpyridin-2-yl derivatives. The size of the substituent at C4 and the 3-nitro→3-amino switch have a profound impact on activity against SA. Amino derivatives are less prone toward hydrolysis of the lactone ring. This fact represents a determinant of activity against SA shared by novobiocin and 3-amino derivatives from this study.

3.2. QSAR. 3.2.1. Model Validation. Selected statistical and machine learning methods were applied to the results of the antimicrobial activity screening against CA and SA. LOO external validation results are given in Table 3:

Statistical analysis of correspondence between RF and either GA/1NN or GA/SVM has been made. The Man-Whitney U test with 0.05 significance did not reveal any statistically significant differences between accuracies of analyzed pairs. Still, in comparison to RF, the other two methods are characterized by a higher CV% (i.e., these methods are less stable). In the GA/SVM case this finding was confirmed by Levene's test of variance homogeneity with 0.05 significance. This result is probably caused by the need for parallel chromosome selection and parameter tuning during the training phase. This means that the GA/SVM training strategy not only increases demands for processing time but also introduces the instability of prediction accuracy.

In the CA case, an additional statistical test has been made. The proportion of molecules that are inactive against at least one CA strain is quite small (8/33). If one assigns measurable activity against CA to all selected coumarins, accuracy is expected to be 75.75%. In order to test the prediction accuracy of RF and GA/1NN against the "all are active" approach, a 2×2 χ^2 test was used. Although accuracy averages of the RF and GA/1NN approach are higher than the "all are active" accuracy expectation, both approaches failed to prove statistically significant improvement of accuracy over the "all are active" approach ($p < 0.05$)! This finding calls for further improvement of the RF and GA/1NN approaches. In accordance with Somorjai's suggestions, predictive performance could be improved by application of two or more predictors at the same time.[30] In this situation we simply decided to assign new candidates to the active class only if both GA/1NN and RF predictors attribute them to that class. This way, the probability of assigning a candidate to the inactive class if it is truly active in the CA case falls from 12.40% and 8.00%, respectively, to 0.99%, under assumption of method independence. The false positive error rate also decreases from 52.50% and 61.25%, respectively, to 32.16% in this way. However, false rejection of truly active compounds could happen when two methods predict different activity classes for the same candidate, which could be considered as the deficiency of this strategy.

Table 4. Determinants of Antimicrobial Activity of Analyzed Coumarins against Selected Microorganisms[b]

	microorganism			
	SA		CA	
	GA/SVM[a]	RF[a]	GA/1NN[a]	RF[a]
descriptor rank	biological determinants			
1	MCS4Max	**MC9**	D	TV11
2	MC4	**MCS4Sum**	TVS4LA	TV10
3	SAS	TV8	BO4S4	MR
4	**MCS4Sum**	MC7	MC3	**BO78**
5	**BO12**	TV4	BOS3Max	MC4
6	MCS3Max	**BO12**	TV6	R12
7	MC10	R12	**BO78**	R211
8	TV2	R4S4	R34	BO56
9	BO510	TVS4LA	R312	BO89
10	**MC9**	TV9	TVS3Min	MC5

[a] Method. [b] BOS3Max − refers to the higher BO value among two values corresponding to the S3; D − dipole moment; MCS3Max − refers to the O atom belonging to the S3 substituent characterized with the lower MC value; MCS4Max − refers to the maximum MC value corresponding to the S4; MCS4Sum − refers to the sum of the MC values corresponding to the S4; TVS3Min − refers to the O atom belonging to the S3 substituent characterized with the lower TV value; TVS4LA − refers to the TV value of an atom that links S4 to the position 4 of the coumarin ring. The descriptors given on both lists that correspond to a single microorganism are written in bold letters. Complete ranking lists are available on request.

3.2.2. Antimicrobial Activity Dependence on Values of Selected Descriptors. The highest-ranking descriptors of antimicrobial activity against SA and CA selected after applying the approaches described in the previous text are given in Table 4.

Applied QSAR methods enable descriptor ranking and selection. However, descriptor ranking is highly dependent on the selected method. In order to find method independent relevant descriptors, only the ones given on both lists (Table 4) that correspond to a single microorganism were selected. The number of determinants is also lowered in this way and that makes interpretation of the results easier. Figures 1 and 2 represent activities of analyzed coumarins against selected microorganisms in the space defined by two determinants obtained in the described fashion.

First of all, according to Table 4 there are three relevant, i.e., independent, determinants of antimicrobial activity against SA: MCS4Sum, MC9, and BO12. The first two determinants that were selected for graphical representation are given in Figure 1. In the selected MC9/MCS4Sum representation active derivatives are almost perfectly separated from inactive ones by a linear separation line (Figure 1). This figure enables simple analysis of relationship between selected determinants and the activity. According to Figure 1 antimicrobial activity is inversely proportional to MCS4Sum. On the other hand, as the MC9 value approaches zero, antimicrobial activity increases. The third relevant descriptor is BO12. In order to determine its influence on antimicrobial activity univariate logistic regression based on this descriptor was performed. According to the univariate logistic model antimicrobial activity against SA is significantly correlated ($p < 0.05$) with BO12.

Molecules active against SA are characterized by the negative sum of partial atomic charges corresponding to the S4 (Figure 1). The importance of localization of negative

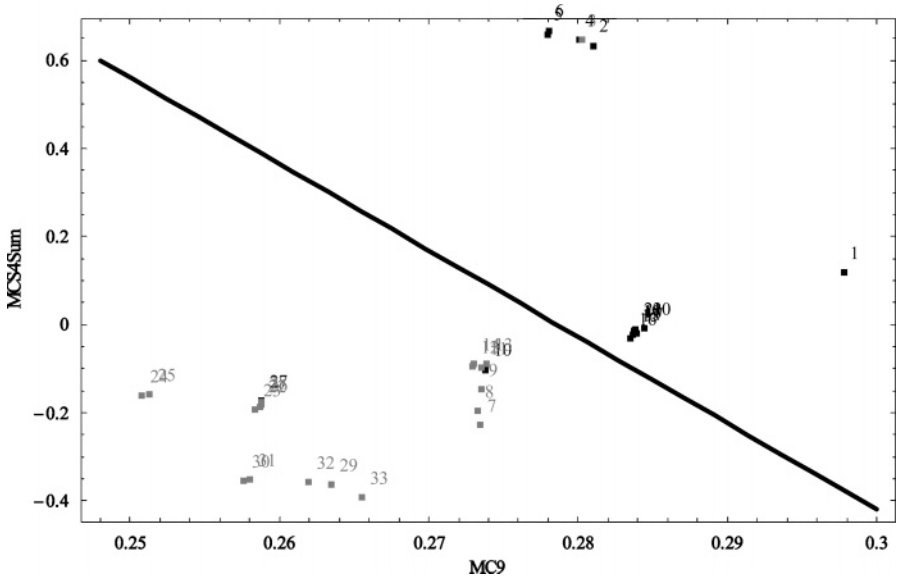


Figure 1. Antimicrobial activity of analyzed coumarins against SA. For the representation of molecules space defined by partial atomic charge of atom C9 (MC9) and sum of partial atomic charges corresponding to the S4 (MCS4Sum) has been chosen (Table 4). Active compounds are represented by gray dots, while inactive compounds are represented by black dots.

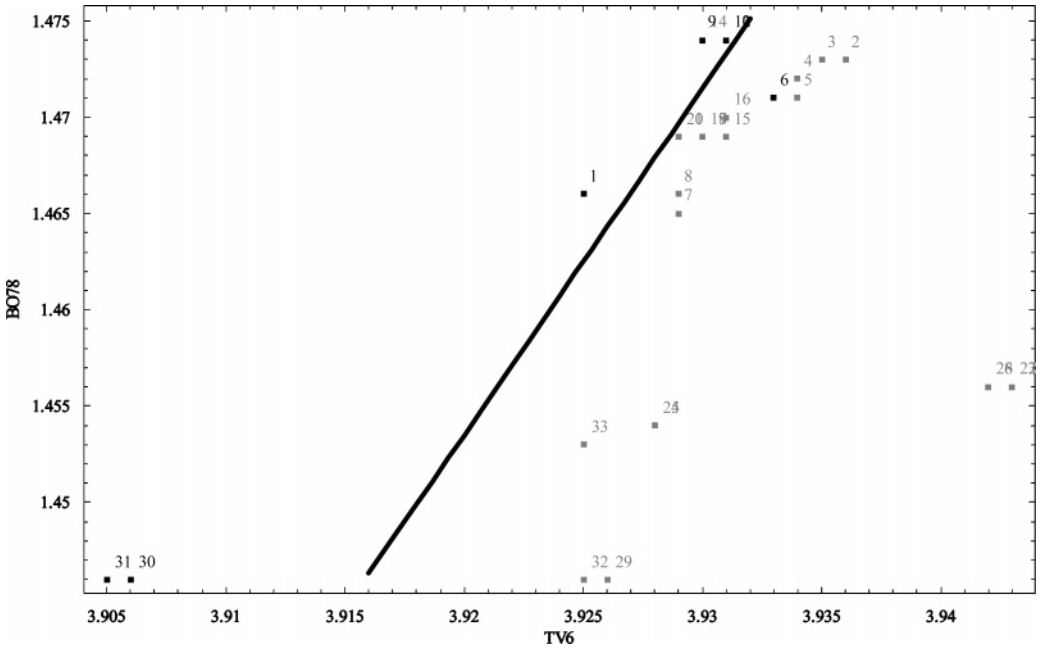


Figure 2. Antimicrobial activity of analyzed coumarins against CA. For the representation of molecules space defined by total atomic valence of the atom C6 (TV6) and order of bond between atoms C7 and C8 (BO78) has been chosen (Table 4). Active compounds are represented by gray dots, while inactive compounds are represented by black dots.

charge at S4 is probably related to the dipole−dipole type of interaction of coumarin derivatives with target molecules. High values of BO12 and low values of MC9 needed for activity indicate the importance of the O1−C2 bond. It has been shown that coumarins with electron-withdrawing substituents attached to C3 and/or C4 atoms are prone to hydrolysis of this bond.[34] Direct evidence of the tendency toward hydrolysis is behavior of BO12. MC9 is probably indirectly connected to this phenomenon since the C9 atom is attached to O1. It is also noteworthy that S3 descriptors are almost absent from the top ranking descriptors list, although corresponding SAR implied a significant role of this substituent. It seems that the S3 has an indirect influence on antimicrobial activity against SA. One possible explanation lies in the fact that a switch from a nitro to amino substituent at the position C3 represents a switch from electron-withdrawing to electron-donating substitution. This type of substitution stabilizes the O1−C2 bond, and it is directly reflected in the BO12 increase.

First of all, according to Table 4 there is only one independent descriptor—BO78 of antimicrobial activity against CA. However, descriptor TV6 has been placed at the sixth and 11th positions of GA/1NN and RF rankings, respectively, and therefore it was selected for generation of Figure 2. Acceptable linear separation of molecules has been achieved this way. Relatively high values of TV6 are needed for measurable antimicrobial activity against CA, which is especially pronounced when BO78 values are high.

Selection of these descriptors is obviously connected to the aromaticity of the benzene ring. If BO78 reaches its

optimal value of 1.5 and TV6 keeps its value as close as possible to its optimal value of 4, a molecule has a good chance to be active against CA. The product of TV6 and BO78 corresponding to active molecules approximately equals 5.7 or 5.8. This rule holds even for molecules with low BO78 values, i.e., 3-nitro-4-(pyridin-2-ylimino)chromen-2-ol derivatives (compounds 29−33). High aromaticity, i.e., electron delocalization on the benzene ring, especially in the C6−C7−C8 fragment, could play an important role in some kind of hydrophobic interaction with the active site of the molecular target in the case of CA.

R12 has been placed at the sixth position of the RF list and the 17th position of the GA/1NN list. Its role also could be related to hydrolytic instability of analyzed coumarins since an increase in R12 values is connected to bond cleavage. Unfortunately, a statistically significant univariate logistic model of its impact on antimicrobial activity against CA could not be established. This result indicates nonlinear dependence of antimicrobial activity against CA on R12 and/or multivariate behavior of antimicrobial activity against CA. A similar conclusion could be drawn about D, which was given a no 1. ranking on the GA/1NN list. This descriptor is related to charge distribution across the coumarin backbone. It is expected that low values of D correspond with a high activity, since they correlate with a high degree of covalent nature of the O1−C2 bond (in the SA case this has been statistically confirmed). Unfortunately, this descriptor did not yield a statistically significant univariate logistic model of antimicrobial activity against CA.

3.3. SAR and QSAR Results Integration. The lack of activity of compound 1 points to the importance of a heterocyclic and/or electron-withdrawing substitution at the position 4 of the coumarin (chromen) ring. In the SA case this finding is confirmed by the mandatory demand for activity: negative values of MCS4Sum. In the CA case the lack of activity of compounds 12 and 13 and their counterparts, molecules 30 and 31, remains unexplained. However, this finding accompanied by the lack of activity of compound 1 implies that S4 could be engaged in some kind of steric interaction with the target.

Molecules 29−33, in which nonplanar PM3 geometry is stabilized by the intramolecular H-bonds, are characterized by substantial activity against SA. This fact proves that the coumarin ring planarity is not the essential property that determines antimicrobial activity against SA. Even in the CA case some of the members of this group show measurable activity. Therefore this conclusion stands for CA as well. These molecules, especially compound 33, exclude the possibility of DNA gyrase inhibition since they lack an oxo group at the position C2 that is a necessary prerequisite for coumarin docking.[17]

Since the ClogP is not present among determinants of antimicrobial action against any of the analyzed microorganisms, the interaction of selected coumarins with lipids is also of a minor importance. Still, independency and relevance of BO78 and TV6 imply some kind of hydrophobic interaction in the CA case. This conclusion cannot be withdrawn for the SA case.

On the other hand, high antimicrobial activity of 3-amino derivatives (compounds 22−28) against both microbes indicates the significance of electron-donating versus -withdrawing properties of the S3 substituent. This finding could be explained by stabilization of the O1−C2 bond. According to Table 4 three descriptors could be found at least in one instance on lists corresponding to each of the analyzed microorganisms. These descriptors are TVS4LA, R12, and MC4. Properties of the O1−C2 bond and atomic charge of the C4 atom are obviously common characteristics of antimicrobial action of analyzed coumarins. While the background of shared TVS4LA and MC4 importance could be related to chemical activity of the atom C4, O1−C2 hydrolytic instability could be the hallmark of analyzed coumarins.

Based on the SAR and QSAR results differences among antimicrobial action of selected coumarins against CA and SA are obvious (Table 1 and 4, Figures 1 and 2). These differences indicate that different molecular targets mediate antimicrobial activity in the SA and CA case. Importance of the O1−C2 bond properties represents a common feature that determines activity against both of these microorganisms. A shared MC4 determinant could be related to Michaels' addition of a cysteine side-chain thiol group of a potential protein target molecule to the C4 atom, as suggested by D. Mares.[14] In the SA case the significance of MCS4Sum, which indicates the importance of the electron-withdrawing properties of the S4, supports this hypothesis. Moreover, evidence from chemical experiments show high reactivity of the C4 toward an electrophilic attack.[6]

3.4. New Candidates. In the previous section, along with the analysis of antimicrobial determinants, qualitative and quantitative directives for the design of potent derivatives have been given. SARs from the previous text provide a rough definition of the space of possible coumarin substitutions that ensure activity against SA and CA. QSAR results (Figures 1 and 2) provide details about the relationship between activity and calculated descriptors, and, in this way they provide the means for narrowing the set of possible candidates.

Elimination of the S3 substituent seemed attractive since it was shown that electron-withdrawing substituents could cause the instability of the lactone structure. Even more, some closely related derivatives without an S3 substitution possess certain pharmacological activity.[35] Nevertheless, in order to avoid unsupported modeling in the space of untested substitutions or extrapolation the modeling process was limited to compounds substituted at the C4 and C3 positions. SAR analysis has shown that small heterocyclic substituents directly linked to the position C4 possess considerable activity toward both microorganisms. Besides, existence of a 3-amino group improves activity against SA and probably stabilizes the lactone ring.

QSAR analysis reveals the need for a negative cumulative charge placed at the S4 in the SA case. This demand and the demand for small heterocyclic substituents at the C4 position are fulfilled by the selection of N- or O-containing small rings. Considerable improvement of antimicrobial potential, especially against SA, is expected to be achieved by the introduction of more than one heteroatom. Nonaromatic heterocycles placed at the C4 position are also included among the candidates. Although the initial set of compounds does not contain molecules with this type of substituent, descriptor values that correspond to molecules with this type of S4 fall in the ranges covered by molecules from the initial set. This way extrapolation has been avoided. In the CA case,

Table 5. MIC of Selected Compounds

compound	*S. aureus* ATCC 29213[a] MIC/μg/mL	*C. albicans* ATCC 90028[a] MIC/μg/mL
8	>256	1024
23	>256	64
C51	>256	256
C52	>256	256

[a] Strain.

selected substitutions should not impair aromaticity of the benzene ring. Finally, BO12 values should be close to 1 in order to eliminate significant O1−C2 bond hydrolysis. Due to the space requirements, candidate molecules that satisfy given demands are placed at the end of Table 1. Structure activity predictions are also given. Activity prediction for each candidate has been repeated ten times with an altered random seed. The activity predictions given in Table 1 are obtained by a majority voting across these repetitions.

4-Morpholino coumarin derivatives (compounds C51 and C52) represent candidates with the most promising activity against both microorganisms (Table 1). In these cases all predictions were positive for each value of random seed and for each prediction method. Besides these molecules, the rest of the candidates also show potential activity at least against CA. However, some of these candidates possess some undesirable characteristics like a complex synthesis. In order to reduce the margin of error and in order to avoid synthesis and microbiological testing related problems, 4-morpholino derivatives (compounds C51, C52) were selected for final MIC determination (Table 5).

Table 5 shows that SAR and QSAR analyses did not yield considerable improvement of antimicrobial activity against SA. However, MIC values even of the most potent derivatives (compounds 8 and 23) from the initial set were above the standard measurement range. On the other hand, in the CA case designed molecules C51 and C52 are placed among the most potent derivatives from the initial set according to the measured MIC values. These values are still quite high.

The most important limitation of the presented QSAR study is the binary nature of the antimicrobial activity measure provided for the initial set. This type of response variable allows only distinction between active and inactive molecules, while the activity improvement over the limits covered by initial microbiological activity cannot be assured by this approach. In order to do so one needs continuous measures like MIC in this case. Within these limitations the applied approach yielded acceptable performance.

4. CONCLUSIONS

The presented study revealed determinants of 3-nitrocumarins' and related compounds' antimicrobial activity against C. *albicans* and S. *aureus*. In the CA case relevant descriptors are C7−C8 bond order and C6 atomic valence. Values of these descriptors around 1.5 and 4, respectively, are characteristic of derivatives that are active against CA. This finding indicates that the C6−C7−C8 fragment and its aromatic character play an important role in antimicrobial activity against CA. In the SA case MCS4Sum, MC9 and BO12 represent relevant determinants. Coumarin derivatives should possess a negative value of the S4 cumulative partial atomic charge in order to be active against SA. The partial charge of C9, and especially the O1−C2 bond order, points to a significant role of O1−C2 bond properties. Hydrolysis of this bond could diminish the effect of analyzed derivatives on SA. Length of the same bond as well as the molecular dipole moment could be related to this phenomenon as well. Although these descriptors take high rankings in the CA case, their role and the role of the O1−C2 bond in antimicrobial activity against CA have not been revealed.

Guidelines provided by Figures 1 and 2 and validated class prediction models provided the means for a design of new derivatives with antimicrobial activity against these two microorganisms. Among selected molecules derivatives with a morpholino substituent placed at the position C4 were the most promising in terms of predicted antimicrobial activity against both microorganisms. Experimental evaluation of activity of these derivatives proved that activity of candidates is comparable to activity of representatives of the initial set. Still, MIC values of all molecules tested against both microorganisms are still too high. According to these results continuation of work on the antimicrobial activity improvement of this class of compounds against SA has no justification. On the other hand, improvement of the activity of these compounds against CA in order to achieve acceptable MIC seems feasible. QSAR could be helpful in this instance, but MIC should be used as a response variable.

Finally, comparison of RF and two GA based methods revealed that RF and the other two methods provide similar average accuracies. However, RF with predefined and fixed user-defined values proved to be more stable toward random seed alternation.

Abbreviations: ATCC − American type culture collection; ATP − adenosine triphosphate; BO − bond order; CA − *Candida albicans*; CLSI − clinical and laboratory standards institute; CV − coefficient of variation; DFT − density functional theory; DIZ − diameter of inhibition zone; GA − genetic algorithm; kNN − k nearest neighbor; LOO − leave-one-out; MC − Mulliken charges; MIC − minimal inhibitory concentration; MR − molar refractivity; Mr − molecular (relative) mass; PI-PLC − phosphatidylinositol phospholipase C; QSAR − quantitative structure−activity relationship; SA − *Staphylococcus aureus*; SAR − structure−activity relationship; SVM − support vector machines; TV − total valence.

ACKNOWLEDGMENT

The authors would like to thank to the following individuals for useful discussions and technical support: David Meyer, Vienna University of Economics and Business Administration; Egon Willighagen, Radboud University Nijmegen; Andy Liaw and Vladimir Svetnik, Merck Research Laboratories at New Jersey; Holger Frölich, Center for Bioinformatics at Tübingen; Željka Gjuranović and Valerije Vrček, Faculty of Pharmacy and Biochemistry, University of Zagreb; Lidija Žele-Starčević, KBC Zagreb; and Ivaylo Elenkov, GSK Research Centre Zagreb. The authors would also like to thank to the reviewers for constructive criticism and the Croatian Ministry of Science for a financial support through project No. 0006541 (M.M.-Š.).

REFERENCES AND NOTES

(1) Rodighiero, G.; Antonello, C. Sintesi di Alcuni Derivati 3-Ammino Cumarinici e Prime Notizie Sulle Loro Proprietá Antibatteriche. *Boll. Chim. Farm.* **1958**, *97*, 592−601.

(2) Kulkarni, M. V.; Patil, V. D. Studies on Coumarins I. *Arch. Pharm.* **1981**, *314*, 708−711.

(3) Kulkarni, M. V.; Patil, V. D. Studies on Coumarins II. *Arch. Pharm.* **1983**, *316*, 15−21.

(4) Reusser, F.; Dolak, L. A. Novenamine is the Active Moiety in Novobiocine. *J. Antibiot.* **1986**, *39*, 272−274.

(5) Althaus, I. W.; Dolak, L.; Reusser, F. Coumarins as Inhibitors of Bacterial DNA Gyrase. *J. Antibiot.* **1988**, *41*, 373−376.

(6) Govori, S.; Rapić, V.; Leci, O.; Čačić, M.; Tabaković, I. Synthesis and Reactions of Some 4-Heteroaryl-3-Nitrocoumarins. *J. Heterocycl. Chem.* **1996**, *33*, 351−354.

(7) Hoult, J. R. S.; Paya, M. Pharmacological and Biochemical Actions of Simple Coumarins: Natural Products with Therapeutic Potential. *Gen. Pharmacol.* **1996**, *27*, 713−722.

(8) Laurin, P.; Ferroud, D.; Klich, M.; Dupuis-Hamelin, C.; Mauvais, P.; Lassaigne, P.; Bonnefoy, A.; Musicki, B. Synthesis and *In Vitro* Evaluation of Novel Highly Potent Coumarin Inhibitors of Gyrase B. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2079−2084.

(9) Vieira, P. C.; Mafezoli, J.; Pupo, M. T.; Fernandes, J. B.; da Silva, M. F. G. F.; de Albuquerque, S.; Pavao, F. Strategies for the Isolation and Identification of Trypanocidal Compounds from the Rutales. *Pure Appl. Chem.* **2001**, *73*, 617−622.

(10) Al-Haiza, M. A.; Mostafa, M. S.; El-Kady, M. Y. Synthesis and Biological Evaluation of Some New Coumarin Derivatives. *Molecules* **2003**, *8*, 275−286.

(11) Galm, U.; Dessoy, M. A.; Schmidt, J.; Wessjohann, L. A.; Heide, L. In Vitro and In Vivo Production of New Aminocoumarins by a Combined Biochemical, Genetic and Synthetic Approach. *Chem. Biol.* **2004**, *11*, 173−183.

(12) Schmutz, E.; Mühlenweg, A.; Li, S.; Heide, L. Resistence Genes of Aminocoumarin Producers: Two Type II Topoisomerase Genes Confer Resistence against Coumermycin A_1 and Clorobiocin. *Antimicrob. Agents Chemother.* **2003**, *47*, 869−877.

(13) Heinrich, M.; Barnes, J.; Gibbons, S.; Williamson, E. M. *Fundamentals of Pharmacognosy and Phytotherapy*; Churchill Livingstone: Edinburgh, U.K., 2004; pp 73−75.

(14) Mares, D. Antimicrobial Activity of Protoanemonin, a Lactone from Ranunculaceous Plants. *Mycopathologia* **1987**, *98*, 133−140.

(15) Škrbo, A. Newly Synthesized Coumarin Derivatives: Structure-Activity Relationship, Ph.D. Thesis, University of Sarajevo, Sarajevo, BiH, 1994.

(16) Tisi, R.; Coccetti, P.; Banfi, S.; Martegani, E. 3-Nitrocoumarin is an Efficient Inhibitor of Budding Yeast Phospholipase-C. *Cell Biochem. Funct.* **2001**, *19*, 229−235.

(17) Lewis, R. J.; Singh, O. M. P.; Smith, C. V.; Skarzynski, T.; Maxwell, A.; Wonacott, A. J.; Wigley, D. B. The Nature of Inhibition of DNA Gyrase by Coumarins and the Cyclothialidines Revealed by X-Ray Crystallography. *EMBO J.* **1996**, *15*, 1412−1420.

(18) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347−1363.

(19) Chou, J. T.; Jurs, P. C. Computer-assisted Computation of Partition Coefficient from Molecular Structures Using Fragment Constants. *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 172−178.

(20) Ghose, A.; Pritchett, A.; Crippen, G. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships: Modeling Hydrophobic Interactions. *J. Comput. Chem.* **1988**, *9*, 80−90.

(21) Bodor, N.; Gabanyi, Z.; Wong, C. A New Method for Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783−3786.

(22) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5−32.

(23) Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J. C.; Sheridan, R. P.; Feuston, B. P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947−1958.

(24) Fröhlich, H.; Chapelle, O.; Schölkopf, B. Feature Selection for Support Vector Machines Using Genetic Algorithms. *Int. J. Artif. Intell. Tool.* **2004**, *13*, 791−800.

(25) Deutsch, J. M. Evolutionary Algorithms for Finding Optimal Gene Sets in Microarray Prediction. *Bioinformatics* **2003**, *19*, 45−52.

(26) Debeljak, Ž.; Marohnić, V.; Srečnik, G.; Medić-Šarić, M. Novel Approach to Evolutionary Neural Network Based Descriptor Selection and QSAR Model Development. *J. Comput.-Aided. Mol. Des.* **2005**, *19*, 835−855.

(27) Baumann, K. Cross-Validation as the Objective Function for Variable-Selection Techniques. *Trends Anal. Chem.* **2003**, *22*, 395.

(28) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms - Part 1. Concepts, Properties and Context. *Chemom. Intell. Lab. Syst.* **1993**, *19*, 1−33.

(29) Lucasius, C. B.; Kateman, G. Understanding and Using Genetic Algorithms - Part 2. Representation, Configuration and Hybridization. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 99−145.

(30) Somorjai, R. L.; Dolenko, B.; Baumgartner, R. Class Prediction and Discovery Using Gene Microarray and Proteomics Mass Spectroscopy Data : Curses, Caveats, Cautions. *Bioinformatics* **2003**, *19*, 1484−1491.

(31) *R: A Language and Environment for Statistical Computing*; R Development Core Team: Vienna, Austria, 2005.

(32) Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing Model Fit by Cross-Validation. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 579.

(33) Savel'ev, V. L.; Artamonova, O. S.; Troitskaya, V. S.; Vinokurov, V. G.; Zagorevskii, V. A. Investigations of pyrans and related compounds. *Chem. Heterocycl. Compd.* **1973**, *9*, 816−820.

(34) Bowden, K.; Hanson, M. J.; Taylor, G. R. Reactions of Carbonyl Compounds in Basic Solutions. Part I. The Alkaline Ring Fission of Coumarins. *J. Chem. Soc. B* **1968**, 174−177.

(35) Roma, G.; Di Braccio, M.; Carrieri, A.; Grossi, G.; Leoncini, G.; Signorello, M. G.; Carotti, A. Coumarin, Chromone, and 4(3*H*)-pyrimidinone Novel Bicyclic and Tricyclic Derivatives as Antiplatelet Agents: Synthesis, Biological Evaluation, and Comparative Molecular Field Analysis. *Bioorg. Med. Chem.* **2003**, *11*, 123−138.

CI600473Z