Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 1651–1657
Bioorganic & Medicinal Chemistry Letters

QSAR for anti-RNA-virus activity, synthesis, and assay of anti-RSV carbonucleosides given a unified representation of spectral moments, quadratic, and topologic indices

Humberto González-Díaz,[a,*,†] Maykel Cruz-Monteagudo,[b] Dolores Viña,[a] Lourdes Santana,[a] Eugenio Uriarte[a] and Erik De Clercq[c]

[a]*Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Spain*
[b]*Applied Chemistry Research Centre, Central University of Las Villas, 54830, Cuba*
[c]*Rega Institute for Medical Research, Katholieke Universiteit Leuven, B-3000 Leuven, Belgium*

Received 22 November 2004; revised 18 January 2005; accepted 20 January 2005

Abstract—The unified representation of spectral moments, classic topologic indices, quadratic indices, and stochastic molecular descriptors show that all these molecular descriptors lie within the same family. Consequently, the same prior probability for a successful quantitative-structure–activity-relationship (QSAR) may be expected irrespective of which indices are selected. Herein, we used stochastic spectral moments as molecular descriptors to seek a QSAR using a database of 221 bioactive compounds previously tested against diverse RNA-viruses and 402 nonactive ones. The QSAR model thus obtained correctly classifies 90.9% of compounds in training. The model also correctly classifies a total of 87.9% of 207 compounds on additional external predicting series, 73 of them having anti-RNA-virus activity and 134 nonactive ones. In addition, all compounds were regrouped into five different subsets for leave-group-out studies: (1) anti-influenza, (2) anti-picornavirus, (3) anti-paramyxovirus, (4) anti-RSV/anti-influenza, and (5) broad range anti-RNA-virus activity. The model has retained overall accuracies of about 90% on these studies validating model robustness. Finally, we exemplify the practical use of the model with the discovery of compounds **124** and **128**. These compounds presented MIC_{50} values = 3.2 and 8 μg/mL against respiratory syncytial virus (RSV) respectively. Both compounds also have low cytotoxicity expressed by their Minimal Cytotoxic Concentrations >400 μg/mL for HeLa cells. The present approach represents an effort toward a formalization and application of molecular indices in bioorganic and medicinal chemistry.

Quantitative-structure–activity relationships (QSAR) have emerged due to the interest of researchers worldwide on finding timely and rational ways for the discovery of new drug-like compounds including antibacterial, anti-parasitic, and anti-viral compounds.[1–8] The QSAR directed discovery of anti-virals active against RNA viruses has become a pressing problem as the result of the relative widespread use of a few commercial drugs causing the emergence of anti-viral-resistant pathogens and the large amount of orphan viral diseases. The application of QSAR techniques and molecular descriptors for anti-viral discovery has relied mainly on anti-HIV-viral drugs. As a consequence, the field of QSAR devoted to other anti-RNA-viral compounds remains practically unexplored.[9] In recent years, along with this discovery process we have explored a large number of nucleosides analogues, which have been successfully designed and synthesized.[10] A panoply of defined molecular descriptors imposes the necessity of unified theories for the systemization of molecular indices, which may guide authors in their selection.

Molecular descriptors can be grouped in families in order to facilitate their study. Almost all of the most popular molecular descriptors can be expressed by means of vector–Matrix–vector ($\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^{T}$) representations. For instance, the first molecular descriptor defined in a chemical context, the Wiener index W (Eq. 1), is a

Keywords: QSAR; Markov models; Spectral moments; Topologic indices; Quadratic indices; Antiviral activity; Carbonucleoside synthesis.
* Corresponding author at present address: Department of Organic Chemistry, Faculty of Pharmacy, USC, Campus Sur. 15782, Santiago de Compostela, Spain. Tel.: +34 981 563100 (14938); fax: +34 981 594 912; e-mail: humbertogd@vodafone.es
† Leaving from CBQ, UCLV, 54830, Cuba.

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.01.047

quadratic form. In addition, there are several other classic Zagreb indices M_1 (Eq. 2) and M_2 (Eq. 3), Harary number H (Eq. 4), Randic invariant χ (Eq. 5), valence connectivity index χ^v (Eq. 6), the Balaban index J (Eq. 7), the Moreau–Boroto autocorrelation ATS_d (Eq. 8).[11–13] More recently other topologic indices based on quadratic forms, the so-called quadratic indices $q_k(X)$ Eq. 9 have been introduced by Marrero et al.[14]

$$W = \frac{1}{2}(\mathbf{u}\cdot\mathbf{D}\cdot\mathbf{u}^{\mathrm{T}}) \tag{1}$$

$$M_1 = (\mathbf{v}\cdot\mathbf{A}\cdot\mathbf{u}^{\mathrm{T}}) \tag{2}$$

$$M_2 = \frac{1}{2}(\mathbf{v}\cdot\mathbf{A}\cdot\mathbf{v}^{\mathrm{T}}) \tag{3}$$

$$H = \frac{1}{2}(\mathbf{u}\cdot\mathbf{D}^{-k}\cdot\mathbf{u}^{\mathrm{T}}) \tag{4}$$

$$\chi = \mathbf{v}'\cdot\mathbf{A}\cdot\mathbf{v}'^{\mathrm{T}} \tag{5}$$

$$\chi^v = \mathbf{v}''\cdot\mathbf{A}\cdot\mathbf{v}''^{\mathrm{T}} \tag{6}$$

$$J = \frac{1}{2}\cdot C\cdot(\mathbf{d}'\cdot\mathbf{A}\cdot\mathbf{d}'^{\mathrm{T}}) \tag{7}$$

$$\mathrm{ATS} = \mathbf{w}^m\cdot\mathbf{B}\cdot\mathbf{w}^{\mathrm{T}} \tag{8}$$

$$q_k(X) = \mathbf{w}\cdot\mathbf{M}\cdot\mathbf{w}^{\mathrm{T}} \tag{9}$$

All the vectors and matrices used in expressions 1–9 have been exhaustively explained in the literature, see therein for details.[15]

On the other hand, several studies have made use of the concept of molecular descriptors based on spectral moments.[16–34,21,35–39] This group of molecular descriptors has classically been considered as a different group with respect to classic topologic indices. Spectral moments have presented several applications in different contexts such as polymer sciences, solid-phase chemistry, and theoretic chemistry. In QSAR and bioorganic chemistry several applications have been reported by Gónzalez et al.[30–34] Morales et al.[34,21] Cabrera-Pérez et al.[35–37] Molina et al.[38] Estrada and Peña,[39] and others. All these spectral moment indices used in the above mentioned studies and others including the moments of energy $\mu(\mathbf{H})$, the self-return walking counts srwc^k, the spectral moments of bond $\mu(\mathbf{B})$ and bond weighted adjacency matrices $\mu(^{\mathbf{d}}\mathbf{B})$ matrices, the I_3 number, the Kirchhoff number Kf,[13,16–34,21,35–39] and our stochastic moments $^{\mathrm{SR}}\pi_k$,[40–45] have to be represented as the trace (Tr) of the corresponding matrices and classified (as mentioned above) as a group apart from $\mathbf{v}\cdot\mathbf{M}\cdot\mathbf{v}^{\mathrm{T}}$ forms indices if one follows classic ideas, where atom adjacency (**A**), bond adjacency (**B**), Hückel Hamiltonian (**H**), bond weight diagonal matrix (**W**), Laplacian (**L**), backbone dihedral angles $\mathbf{A}(\varphi, \Psi, \omega)$ and Markov ($^{1}\mathbf{\Pi}$) are well known matrices.[16–34,21,35–39] In particular, our group has worked on a Markov model that use stochastic spectral moments $^{\mathrm{SR}}\pi_k$ as descriptors to encode molecular structure with applications in nucleic acids, proteins, and bioorganic medicinal chemistry research:[40–46]

$$\mathrm{srwc}_k = \mathrm{Tr}(\mathbf{A}^k) \tag{10}$$

$$\mu_k(B) = Tr(\mathbf{B}^k) \tag{11}$$

$$\mu_k(^{d}\mathbf{B}) = \mathrm{Tr}[(^{d}\mathbf{B} + \mathbf{W})^k] \tag{12}$$

$$\mu_k(\mathbf{H}) = \mathrm{Tr}(\mathbf{H}^k) \tag{13}$$

$$Kf = a\cdot\mathrm{Tr}(\mathbf{L}) \tag{14}$$

$$^{\mathrm{SR}}\pi_k = Tr(^{1}\mathbf{\Pi}^k) \tag{15}$$

Interesting steps have been taken toward the unification of all topologic indices in a single framework by means of the vector–matrix–vector ($\mathbf{v}\cdot\mathbf{M}\cdot\mathbf{v}^{\mathrm{T}}$) approach. However, no advances to this promising picture have appeared on the incorporation of spectral moments. The unification of molecular indices as a mathematical representation may facilitate not only its study by researchers worldwide but comprehension of its nature. In the present work, we use the Krocnecker vector **o** in order to represent any spectral moment molecular index as a quadratic form of the corresponding matrix:

$$\mathrm{srwc}_k(\mathbf{A}) = \mathbf{o}\cdot\mathbf{A}\cdot\mathbf{o}^{\mathrm{T}} \tag{16}$$

$$\mu_k(\mathbf{B}) = \mathbf{o}\cdot\mathbf{B}\cdot\mathbf{o}^{\mathrm{T}} \tag{17}$$

$$\mu_k(^{d}\mathbf{B}) = \mathbf{o}\cdot[\mathbf{B} + \mathbf{W}]^k\cdot\mathbf{o}^{\mathrm{T}} \tag{18}$$

$$\mu_k(\mathbf{H}) = \mathbf{o}\cdot\mathbf{H}\cdot\mathbf{o}^{\mathrm{T}} \tag{19}$$

$$Kf = \mathbf{o}\cdot(a\cdot\mathbf{L})\cdot\mathbf{o}^{\mathrm{T}} \tag{20}$$

$$^{\mathrm{SR}}\pi_k = \mathbf{o}\cdot[(^{1}\mathbf{\Pi})^k]\cdot\mathbf{o}^{\mathrm{T}} \tag{21}$$

$$I_3 = \frac{1}{k!}\sum_k^{\infty}\mu_k(\mathbf{A}(\varphi,\Psi,\omega)) = \frac{1}{k!}\sum_k^{\infty}[\mathbf{o}\cdot\mathbf{A}(\varphi,\Psi,\omega)\cdot\mathbf{o}^{\mathrm{T}}] \tag{22}$$

The Kröcnecker elements vector **o** has a simple but dynamic and opportune definition having elements ${}^{m}\delta_{ij} = 1$ for every *j*th column if the element is being multiplied by an element in the main diagonal of the given matrix, and ${}^{m}\delta_{ij} = 0$ otherwise. As can be noted Eqs. 16–22 are $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^{T}$ representations, this fact reveals that spectral moments and stochastic moments may be classified together with several topologic, flexibility, and quadratic indices. That is to say, they all belong to the same family giving a more unified and tractable picture in mathematical chemistry terms. By opposition to classic forms we have named ${}^{SR}\pi_k$ as the stochastic moments $\mathbf{v} \cdot \mathbf{M} \cdot \mathbf{v}^{T}$ forms. Expanding Eq. 21 illustrates more clearly the similarity between classic topologic indices defined in the past, stochastic spectral moments defined by our group in 2002, and Marrero-Ponce et al. quadratic indices.[1,2,11–14,40–49]

$$
{}^{SR}\pi_k = \mathbf{o} \cdot [({}^{1}\mathbf{\Pi})^{k}] \cdot \mathbf{o}^{T} = \begin{bmatrix} {}^{1}o_{ij} & {}^{2}o_{ij} & \cdot & \cdot & {}^{n}o_{ij} \end{bmatrix} \cdot \begin{bmatrix} {}^{1}p_{11} & {}^{1}p_{12} & \cdot & \cdot & {}^{1}p_{1n} \\ {}^{1}p_{21} & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ {}^{1}p_{n1} & \cdot & \cdot & \cdot & {}^{1}p_{nn} \end{bmatrix} \cdot \begin{bmatrix} {}^{1}o_{ij} \\ {}^{2}o_{ij} \\ \cdot \\ \cdot \\ {}^{n}o_{ij} \end{bmatrix} \tag{23}
$$

Expressing the probabilities for the distribution of electrons between the *i*th and the *j*th atom functions of their electronegativities.[50,51] So, the ${}^{SR}\pi_k$ values describe the distribution of electrons to atoms at distance *k* to each other. The definition of different ${}^{1}\mathbf{\Pi}$ matrices with applications in bioorganic chemistry have been largely discussed in the literature by González-Díaz et al.[52–54] The above results demonstrate that classic topologic indices, quadratic indices, spectral moments, and stochastic indices lie together within the same family.

Consequently, we can expect at first instance the same probability of success by selecting one of them for different QSAR studies including antimicrobial agents.[55,56] Thus, the aims of this study were to unify spectral with classic molecular descriptors and develop a new a QSAR model for anti-RNA-viral activity, based on stochastic spectral moments. The linear discriminant analysis (LDA)[57–59] was selected as a simple statistical tool in order to select anti-RNA-virus active compounds from a heterogeneous series. The selection is based on the experience of our group to model biological properties of a heterogeneous series of compounds including carbonucleosides.[60] In this sense, the present study exemplifies the use of the QSAR reported by means of the prediction, synthesis, characterization, and experimental corroboration of the anti-RSV-activity of novel 1, 2-disubstituted carbocyclic analogues of nucleosides.

In order to seek and validate a model for discriminating between anti-RNA-virus and nonactive compounds we have taken the following steps:

(a) The initial data composed by a large number of active and nonactive compounds was collected from the literature and it is presented in this work in Table 1SM and Fig. 1SM as a Supplementary material (SM) file.[61,62]
(b) The molecular structure of all compounds was encoded with the stochastic spectral moments ${}^{SR}\pi_k(\omega)$, which were calculated using the software **BIOMARKS** *version 1.0* (**Bio**chem-informatics **Mark**ovian **S**tudies).[63]
(c) The initial data was split at random into four different sub-series (see Table 1SM and Fig. 1SM):
  - Training series with 221 active compounds.
  - Training series with 402 nonactive compounds.
  - Predicting series with 73 active compounds.
  - Predicting series with 134 nonactive compounds.
(d) The Randič's orthogonalization procedure was applied to each ${}^{SR}\pi_k(\omega)$ variable obtaining the corresponding orthogonal variables ${}^{I}O_k$. This avoids co-linearity among variables and model over-fitting, where the superscript *I* represent the order of importance of the variable assigned after the forward stepwise LDA analysis.[64]
(e) The QSAR–LDA analysis was developed with the software **Statistica 6.0**.[65]
(f) The statistical parameters for different models were compared to decide the model, which better fits the training data. The biological activity was encode by a dummy variable aRNAva (anti-RNA-virus activity), aRNAva = 1 for active compounds and aRNAva = −1 for nonactive ones. The analyzed parameters were Wilk's λ statistic; Mahalanobis squared distance (D^2), Fisher ratio (F), and the p-level (p). We also inspected the percentage of good classification and the proportion between the cases and variables in the equation, or variables to be explored in order to avoid over-fitting or chance correlation.[66]
(g) Model predictability was tested with an external prediction series; those compounds were never used to develop the classification function.[67]
(h) A posterior probability $P(\%)$ was assigned to each compound for scoring its biological activity.[68]
(i) Finally, leave-group-out experiments were carried out to assess the model robustness by checking the stability of all parameters.[69]

As a result of the analysis of co-linearity we detected high regression coefficients among the ${}^{SR}\pi_k$ values for the data. For instance, all regression coefficients among the five calculated descriptors ${}^{SR}\pi_0$ ${}^{SR}\pi_1$ ${}^{SR}\pi_2$ ${}^{SR}\pi_3$ and ${}^{SR}\pi_4$ were higher than 0.9. Subsequently, we carried out a Randič orthogonalization procedure, which results are depicted in Table 2SM of the Supplementary material. Afterwards, we carried out an exhaustive forward stepwise analysis; the best discriminant function we found was the following:

$$
\text{aRNva} = 2.333 \times {}^{1}\text{O}_2 - 1.662 \times {}^{2}\text{O}_1 - 0.651 \tag{24}
$$

$$
N = 623,\ \%T = 90.9,\ \%(-) = 91.5,\ \%(+) = 89.6,
$$
$$
\lambda = 0.50,\ \lambda\text{dif}\% = 25,\ F_{\text{mod}} = 307.0,\ p_{\text{mod}} < 0.001,
$$
$$
F_{\text{last.v.}} = 205.2,\ p_{\text{last.v.}} < 0.001
$$

Where, N is the number of compounds in the model and $\%T$, $\%(+)$, $\%(-)$ are the overall percentage of good classification for anti-RNA-virus and nonactive compounds. Moreover, is the Wilk's statistics and $\Delta\lambda\% = 100 \times (\lambda_s - \lambda_{s-1})/\lambda_s$, and represent the differential decrement in λ in the step s with respect to former step $(s-1)$ in the forward stepwise analysis. Table 2SM also depicts F_m, p_m, and F_l, p_l values, which are equal to the Fisher's ratio and the p-level for the model as a whole (m) and the last variable entered, respectively (see Supplementary material).[31,70]

As depicted in Table 2SM, Supplementary material, the present model with only two variables also presented the higher decrement of $\lambda(\lambda = 0$ ideal separation of groups) with respect to models with 1, 3, 4, and 5. This model also presents a high value for ρ,[6] a parameter, which controls the ratio (number of data points)/(number of fitted parameters), which is expected to be higher than 4 for this kind of analysis. The selected model has shown overall accuracy of 90.9% for training series and 87.9% overall predictability for external predicting series. In the case of nonactive compounds the model correctly classifies 91.5 of nonactive compounds in the training series (see complementary material). On the other hand, the model correctly classified 119 out of 134 of nonactive compounds (88.8%) in the predicting series. With respect to active compounds the model also has shown a good classification of 89.6% in the training series and 86.3% in the predicting series (see Table 1 upper part for summary as well as Table 1SM and Fig. 1SM of the Supplementary material file for details). In the leave-group-out analysis the model showed overall accuracies of 91.8%, 90.8%, 90.45%, 90.1%, and 88.3% after elimination from the starting data of all compounds having anti-influenza, anti-picornavirus, anti-paramyxovirus, both anti-RSV and anti-influenza, or broad range anti-RNA-viral activity, respectively (see Table 1). Accordingly, the robustness of the model for the prediction of anti-RNA-viral compounds, could be assessed after leave-group-procedures. Briefly, after removing different groups of compounds all the parameters of the model lie within the accepted intervals after elimination of different groups of drugs from the model (Table 1 bottom part). The group that caused the higher destabilization of the model when removed was that composed by anti-viral drugs with broader activity against different RNA viruses. This fact is justified because it is not only the group with the largest number of compounds (N out = 340) but also the one with the higher structural diversity. In any case, all the values for the parameters lie within the limits that are classically accepted for LDA–QSAR models in bioorganic medicinal chemistry see for instance Cabrera-Pérez et al. work.[71,72]

Finally, we exemplify the use of the model in practice. In view of the success of the present model, we became interested in using it in our main field of research, 1,2-disustituted carbonucleosides, in which the usual 1,3 substitution pattern of the carbocycle is replaced by a 1,2 pattern. Compounds **124** and **128** were selected among other compounds as examples predicted by LDA-QSAR with high probabilities ($P(\%) = 85.5$ and $\Delta P\% = 74.3$, respectively) and were afterwards synthesized and assayed. It must be noted that compound **123** was predicted as inactive, $P(\%) = 43.0$; however, as it is a synthetic precursor of compound **124** it was also evaluated as additional corroboration to the validity of the model. The three compounds (**123**, **124**, and **128**) where inactive against Vesicular stomatitis virus and Coxsackie virus strain B4. Compound **123** was also inactive against RSV as predicted by the model. Conversely, similar anti-viral activity, MIC_{50} values of 3.2 and 8 μg/mL, were detected against RSV, as compared to 1.92 μg/mL for control anti-viral drug ribavirin, correctly predicted by the model with $P(\%) = 90.1$. The theoretical probabilities and the results of the biological assay of the compounds against Vesicular stomatitisvirus, Cosackie virus B4, Respiratory syncytial virus, as well as cytotoxicity to HeLa cell line were depicted in Table 2.

Compounds **123**, **124**, and **128** were efficiently synthesized from the (±)-*cis*-2-(amino)cyclohexylmethanol following Scheme 1. To obtain the adenine derivative **124**,

Table 1. Accuracy, predictability, and robustness analysis

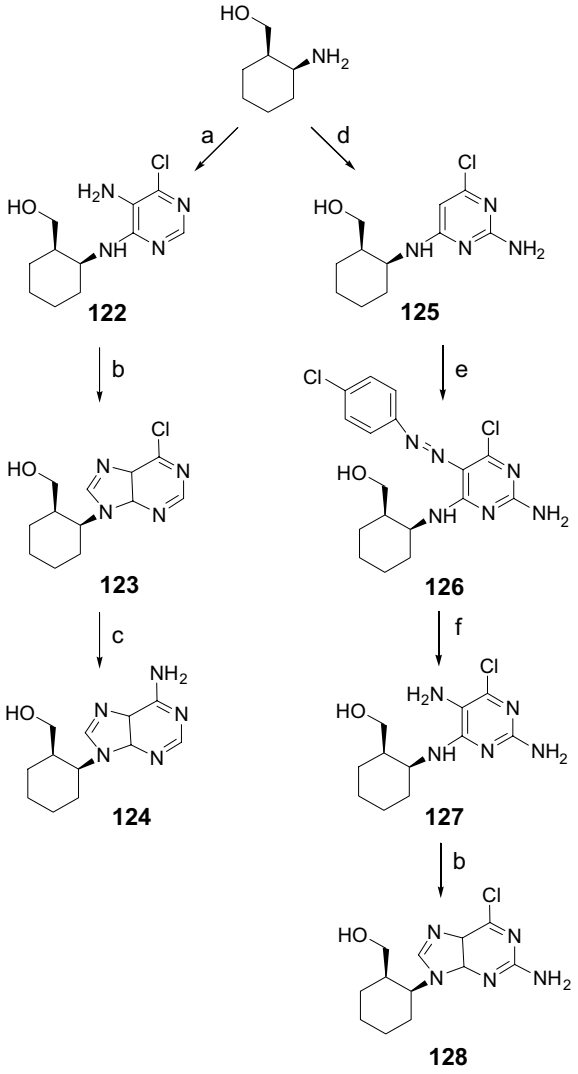
Accuracy and predictability analysis (model with 2 variables)							
	Training series				Predicting series		
	Percent	Active	Non act.		Percent	Active	Non act.
Active	89.6	**198**	23	Active	86.3	**63**	10
Non active	91.5	34	**368**	Non active	88.8	15	**119**
Total	90.9			Total	87.9		

Leave-group-out-robustness-analysis (including together training and predicting series of anti-RNA-virus drugs)					
	Influenza	Picornavirus	Paramyxovirus	RSV and influenza	Broader activity
$\%T$	91.8	90.8	90.45	90.1	88.3
$\%(-)$	91.3	91.3	91.3	91.3	91.3
$\%(+)$	93	89.8	88.9	87.8	73.6
N out	260	235	215	218	340
N in	570	595	615	612	490
λ	0.44	0.49	0.47	0.51	0.75
F	356.02	314.18	343.09	295.53	82.9
P	0.000	0.000	0.000	0.000	0.000

Table 2. Antiviral[a] activity and cytotoxicity[b] of assayed chemical compounds in human epithelial cells (HeLa)

Compound name	Predicted P (%)	Virus			Cytotoxicity
		Vesicular stomatitis	Coxsackie B4	Respiratory syncytial	
123	43.0	>80	>80	>80	400
124	85.5	>80	>80	3.2	400
128	74.3	>200	120	8.0	⩾200
Ribavirin	90.1	16.0	48.0	1.92	⩾400

[a] MIC_{50}: Minimal inhibitory concentration 50 μg/mL.
[b] MCC: Minimal cytotoxicity concentration (μg/mL).



Scheme 1. Reagents and conditions: (a) 5-amino-4,6-dichloropyrimidine, Et_3N, *n*-BuOH, reflux 24 h, 71%; (b) $CH(OEt)_3$, HCl 12 M reflux 12 h, **123**: 71%, **128**: 60%; (c) NH_4OH, reflux 4 h, 99%; (d) 2-amino-4,6-dichloropyrimidine, Et_3N, *n*-BuOH, reflux 24 h, 60%; (e) *p*-chloroaniline, $NaNO_2$, HCl 12 M, 0 °C, 80%; (f) Zn, AcOH, EtOH, reflux 1 h, 30%.

the aminoalcohol was condensed with 5-amino-4,6-dichloropyrimidine to give the substituted diaminopyrimidine **122**, which afforded 6-chloropurine **123** by reaction with ethylorthoformate in acidic medium. The 6-amino derivative **124** was obtained by exchange with ammonium hydroxide. To obtain the 2-amino-6-chloro derivative **128**, the starting aminoalcohol was reacted with 2-amino-4,6-dichloropyrimidine to give **125**. Afterwards a second amino group was introduced at position 5 of the pyrimidine ring by reaction with *p*-chlorobenzenediazonium chloride followed by reduction to afford the compound **127**, which was cyclized with ethylorthoformate to obtain compound **128** (see experimental section in Supplementary material file).[73,74]

In closing, the unification of many molecular descriptors within a single family allows the researcher to begin the study with any one of them without preference. The idea of the unification of different molecular descriptors also facilitates their physicochemical interpretation as in recent bioorganic medicinal chemistry communications by our group.[75,76] Conversely, several classic topologic indices Eqs. 1–22 including Marrero-Ponce's et al. stochastic forms $s_k(X)$ Eq. 25, very similar to our earlier stochastic indices, linear forms $f_k(X)$ Eq. 26, and the above mentioned quadratic forms $q_k(X)$ Eq. 27 lack of direct physical interpretation.[47–49] In contrast, our stochastic forms can be used to derive electrostatic and thermodynamic parameters, see recent works.[75,76]

$$s_k(X) = \mathbf{w} \cdot \mathbf{S}_k \cdot \mathbf{w}^{\mathrm{T}} \tag{25}$$

$$f_k(X) = \mathbf{w} \cdot \mathbf{M} \cdot \mathbf{u}^{\mathrm{T}} \tag{26}$$

$$q_k(X) = \mathbf{w} \cdot \mathbf{M} \cdot \mathbf{w}^{\mathrm{T}} \tag{27}$$

Where, **M** and **S** are the multi-graph and the normalized multi-graph adjacency matrices, and w, u are the electronegativity and the unitary vector.[47–49] In this work, stochastic spectral moments selected priori have been successful for the in silico prediction of anti-RNA-viruses activity. The model has been validated in terms of accuracy, predictability, and robustness to data variation. Taking into consideration that the model was developed with a highly heterogeneous and representative data base of compounds, one can expect a broad range of applicability for it, as exemplified here on the field of 1,2-carbocyclic analogues of nucleosides. The model confirms the utility of stochastic molecular descriptors introduced by González-Díaz et al.[77,78] This model may, as the formers, become a useful tool in bioorganic and medicinal chemistry for the discovery of anti-viral compounds.[79,80]

Acknowledgements

Authors would like to express their gratitude by partial financial support to the University of Santiago de Compostela.

Supplementary data

Enclosed supplementary material depicts names and/or structure together with posterior probabilities of all compounds in train and prediction sets (Table 1SM and Figure 1SM) a summary table for collinearity and forward stepwise analysis (Table 2SM), and the experimental section. Supplementary data associated with this article can be found, in the online version, at 10.1016/j.bmcl.2005.01.047.

References and notes

1. Kier, L. B.; Hall, L. H. *Molecular connectivity in structure–activity analysis*; Research studies press. John Wiley & Sons: Letchworth, England, 1986, pp 225–246.
2. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*; Wiley VCH: Weinheim, Germany, 2000.
3. De Julián-Ortiz, J. V.; Gálvez, J.; Muñoz-Collado, C.; García-Domenech, R.; Gimeno-Cardona, C. *J. Med. Chem.* **1999**, *42*, 3308.
4. García-Domenech, R.; De Julián-Ortiz, J. V. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 445.
5. García-García, A.; Gálvez, J.; De Julián-Ortiz, J. V.; García-Domenech, R.; Muñoz, C.; Guna, R.; Borrás, R. *J. Antimicrob. Chemother.* **2004**, *55*, 65.
6. De Julián-Ortiz, J. V.; De Gregorio Alapont, C.; Ríos-Santamaria, I.; García-Domenech, R.; Gálvez, J. *J. Mol. Graph. Model.* **1998**, *16*, 14.
7. Gozalbes, R.; Brun-Pascaud, M.; García-Domenech, R.; Gálvez, J.; Girard, P. M.; Doucet, J. P.; Derouin, F. *Antimicrob. Agents Chemother.* **2000**, *44*, 2764.
8. Klopman, G.; Wang, S.; Jacobs, M. R.; Bajaksouizian, S.; Edmonds, K.; Ellner, J. J. *Antimicrob. Agents Chemother.* **1993**, *37*, 1799.
9. Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K.; Hansch, C. *Chem. Rev.* **1999**, *99*, 3525.
10. Santana, L.; Teijeira, M.; Teran, C.; Uriarte, E.; Viña, D. *Synthesis* **2001**, 1532.
11. Devillers, J.; Balaban, A. T. Topological Indices and Related Descriptors in QSAR and Drug Design; Amsterdam, 2000; pp 3–41.
12. Kier, L. B.; Hall, L. H. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach Sci.: Amsterdam, 1999, pp 455–489.
13. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley VCH: Weinheim, Germany, 2000.
14. Marrero-Ponce, Y.; González-Díaz, H.; Romero-Zaldivar, V.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2004**, *12*, 5331.
15. Estrada, E. *Chem. Phys. Lett.* **2001**, *336*, 248.
16. Gutman, I.; Rosenfield, V. R. *Theor. Chim. Acta* **1996**, *93*, 191.
17. Estrada, E. *Bioinformatics* **2002**, *18*, 1.
18. Estrada, E. *Chem. Phys. Lett.* **2000**, *319*, 713.
19. González, M. P.; Morales, A. H.; Molina, R. *Polymer* **2004**, *45*, 2773.
20. González, M. P.; Morales, A. H.; González-Díaz, H. *Polymer* **2004**, *45*, 2073.
21. Morales, A. H.; González, M. P.; Rieumont, J. B. *Polymer* **2004**, *45*, 2045.
22. Burdett, J. K.; Lee, S. *J. Am. Chem. Soc.* **1985**, *107*, 3063.
23. Burdett, J. K.; Lee, S. *J. Am. Chem. Soc.* **1985**, *107*, 3050.
24. Lee, S. *Acc. Chem. Res.* **1991**, *24*, 249.
25. Gutman, I. *Theor. Chim. Acta* **1992**, *83*, 313–318.
26. Markovic, S.; Gutman, I. *J. Mol. Struct. Theochem.* **1991**, *81*, 81.
27. Jiang, Y.; Tang, A.; Hoffmann, R. *Theor. Chim. Acta* **1984**, *66*, 183.
28. Karwowski, J.; Bielinska-Waz, D.; Jurkowski, J. *Int. J. Quantum Chem.* **1996**, *60*, 185.
29. Estrada, E.; González-Díaz, H. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 75.
30. González, M. P.; Terán, C. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3077.
31. González, M. P.; Terán, C. *Bioorg. Med. Chem.* **2004**, *12*, 2985.
32. González, M. P.; González-Díaz, H.; Molina, R.; Cabrera-Pérez, M. A.; Ramos de, A. R. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1192.
33. González, M. P.; González-Díaz, H.; Cabrera-Pérez, M. A.; Molina, R. R. *Bioorg. Med. Chem.* **2004**, *12*, 735.
34. González, M. P.; Morales, A. H. *J. Comput. Aided Mol. Des.* **2003**, *10*, 665.
35. Cabrera, M. A.; Bermejo, M.; Gonzalez, M. P.; Ramos, R. *J. Pharm. Sci.* **2004**, *7*, 1701.
36. Cabrera-Pérez, M. A.; García, A. R.; Teruel, C. F.; Álvarez, I. G.; Sanz, M. B. *Eur. J. Pharm. Biopharm.* **2003**, *56*, 197.
37. Cabrera-Pérez, M. A.; González-Díaz, H.; Fernandez, T. C.; Pla-Delfina, J. M.; Bermejo, S. M. *Eur. J. Pharm. Biopharm.* **2002**, *53*, 317.
38. Molina, E.; González-Díaz, H.; González, M. P.; Rodríguez, E.; Uriarte, E. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 515.
39. Estrada, E.; Peña, A. *Bioorg. Med. Chem.* **2000**, *8*, 2755.
40. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernández, S. I.; Morales, A.; Serrano, H. S.; González, J.; Ramos de, A. R. *J. Mol. Model.* **2002**, *8*, 237.
41. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Model.* **2003**, *9*, 395.
42. González-Díaz, H.; Hernández, S. I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
43. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bull. Math. Biol.* **2003**, *65*, 991.
44. González-Díaz, H.; Ramos de, A. R.; Uriarte, E. *Online J. Bioinf.* **2002**, *1*, 83.
45. González-Díaz, H.; Uriarte, R.; Ramos de, A. R. *Bioorg. Med. Chem.* **2005**, *13*, 323.
46. Ramos de, A. R.; González Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
47. Marrero-Ponce, Y. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010.
48. Marrero-Ponce, Y. *Bioorg. Med. Chem.* **2004**, *12*, 6351.
49. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, M.; Mayón-Peréz, M.; García-Sánchez, R. *Bioorg. Med. Chem.* **2004**, *13*, 1293, 2005.
50. Pauling, L. *The Nature of Chemical Bond*; Cornell University: Ithaca, New York, 1939, pp 2–60.
51. Sanderson, R. T. *Polar Covalence*; Academic: New York, 1983.
52. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, I.; Nasco, O.; Uriarte, E.; Castañedo, N. C.;

Cabrera-Pérez, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Toxicol.* **2003**, *16*, 1318.
53. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691.
54. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bioinformatics* **2003**, *19*, 2079.
55. Gozalbes, R.; Gálvez, J.; García-Domenech, R.; Derouin, F. *SAR QSAR Environ. Res.* **1999**, *10*, 47.
56. Gozalbes, R.; Brun-Pascaud, M.; García-Domenech, R.; Gálvez, J.; Girard, P. M.; Doucet, J. P.; Derouin, F. *Antimicrob. Agents Chemother.* **2000**, *44*, 2771.
57. Galvez, J.; Garcia-Domenech, R.; Gomez-Lechon, M. J.; Castell, J. V. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2306.
58. Cronin, M. T. D.; Aynur, A. O.; Dearden, C. J.; Deffy, C. J.; Netzeva, I. T.; Patel, H.; Rowe, H. P.; Schultz, T. W.; Worth, P. A.; Voutzolidis, K.; Schüürmann, G. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 869.
59. Gálvez, J.; García-Domenech, R.; De Julián-Ortiz, J. V.; Soler, R. *J. Chem. Inf. Comput. Sci.* **1994**, *32*, 272.
60. Estrada, E.; Uriarte, E.; Montero, A.; Teijeira, M.; Santana, L.; De Clercq, E. *J. Med. Chem.* **2000**, *43*, 1975.
61. Kleeman, A.; Engel, J.; Kutscher, B.; Reichert, D. *Pharmaceutical Substances*, 4th ed.; George Thieme: Stuttgarti, 2001.
62. Laughlin, A. C.; Tseng, K. C. Burger's Medicinal Chemistry, Wolff, E. M. Ed.; John Wiley & Sons: New York, 1997; Part 4, Vol. 5, pp 561–637.
63. González-Díaz, H.; Molina, R. BIOMARKS version 1.0 (Biochem-informatics Markovian Studies), **2004**, This is and experimental software, contact with the corresponding author humbertogd@vodafone.es.
64. Randič, M. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311.
65. Statsoft, Inc. STATISTICA, version 6.0, 2001.
66. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Chemometric Methods in Molecular Design*; Van Waterbeemd, H., Ed.; Method and Principles in Medicinal Chemistry; Manhnhold, R., Krogsgaard-Larsen, H., Timmerman, H., Eds.; VCH: Weinheim, 1995; Vol. 2, pp 265–282.
67. Kowalski, R. B.; Wold, S. Pattern Recognition in Chemistry. In *Handbook of Statistics*; Krishnaiah, P. R., Kanal, L. N., Eds.; North Holland Publishing Company: Amsterdam, 1982, pp 673–697.
68. Estrada, E.; Uriarte, E. *Curr. Med. Chem.* **2001**, *8*, 1573.
69. Ramos de, A. R.; González-Díaz, H.; Molina, R. R.; Uriarte, E. *Proteins: Struct. Funct. Bioinf.* **2004**, *56*, 715.
70. Gozalbes, R.; Gálvez, J.; Moreno, A.; García-Domenech, R. *J. Pharm. Pharmacol.* **1999**, *51*, 111.
71. Cabrera-Pérez, M. A.; Bermejo, M. *Bioorg. Med. Chem.* **2004**, *22*, 5833.
72. Cabrera-Pérez, M. A.; Bermejo-Sanz, M.; Ramos-Torres, L.; Grau-Ávalos, R.; González, M. P.; González-Díaz, H. *Eur. J. Med. Chem.* **2004**, *39*, 905.
73. Terán, C.; Santana, L.; Uriarte, E.; Viña, D.; De Clercq, E. *Nucleos. Nucleot. Nucl.* **2003**, *22*, 787.
74. Viña, D. Analogos de nucleosidos derivados de ciclopentano y ciclohexano-1,2-disustituidos y de ciclohexeno-1,4-disustituidos. Doctoral Thesis, University of Santiago, Spain, **2003**.
75. González-Díaz, H.; Agüero-Chapin, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castañedo, N. *Bioorg. Med. Chem. Lett.* **2004**, *15*, 551, 2005.
76. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *13*, 1119, 2005.
77. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Polymer* **2004**, *45*, 3845.
78. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de, A. R. *Bull. Math. Biol.* **2004**, *66*, 1285.
79. Gia, O.; Marciani-Magno, S.; González-Díaz, H.; Quezada, E.; Santana, L.; Uriarte, E.; Dalla-Via, L. *Bioorg. Med. Chem.* **2004**, *13*, 809, 2005.
80. De Clercq, E. In *In vitro and ex vivo test systems to rationalize drug design and delivery*; Crommelin, D., Couvreur, P., Duchene, D., Eds.; Editions de Sante: Paris, 1994, pp 108–125.