International Journal of Chemical Modeling
Volume 1, Number 1, pp. 45-62
ISSN: 1941-3955

TURNING SPECTRAL-SAR INTO 3D-QSAR ANALYSIS. APPLICATION ON H^+K^+-ATPASE INHIBITORY ACTIVITY

Mihai V. Putz[1,*], Corina Duda-Seiman[2], Daniel M. Duda-Seiman[3] and Ana-Maria Putz[4]

[1] Laboratory of Computational and Structural Physical Chemistry, Chemistry Department, West University of Timisoara, Timisoara, Romania
[2] Laboratory of Organic Chemistry, Chemistry Department, West University of Timisoara, Timisoara, Romania
[3] University of Medicine and Pharmacy "Victor Babes", Dep. of Medical Ambulatory, Medical Emergencies, Timisoara, Romania
[4] Institute of Chemistry Timisoara of the Romanian Academy, Timisoara, Romania

Abstract

Within the spectral-structure-activity relationship (S-SAR) analysis, the vectorial algebraic version of the consecrated multivariate analysis, the minimum spectral endpoint-connected paths hierarchy is employed to provide the 3-dimensional molecular structure that closely resembles the maximum observed chemical-biological activity across a given series of compounds. The fact that an algebraic correlation factor always exceeds the dispersion based statistically one for the same series of predicted-observed data was here also theoretically proved. As practical illustration of the introduced 3D-Spectral-SAR method the special case of proton pump inhibitors, blocking the acid secretion from parietal cells in the stomach, here belonging to the 4-Indolyl, 2-guanidinothiazole class o derivatives, is exposed and the optimized-by-the-spectral mechanism molecular structure revealed.

Keywords: proton-pump inhibitors, hydrophobicity, polarizability, stericity, spectral paths.

1. Introduction

Inhibition of gastric acid secretion represents the most attractive target in treating acid-related diseases like duodenal and gastric ulcer, gastroesophageal reflux disease (GERD),

* Send correspondence to: Mihai V. Putz; E-mails: mv_putz@yahoo.com, mvputz@cbg.uvt.ro; Tel: +40-0256-592-633; Fax: +40-0256-592-620; Web: http://www.cbg.uvt.ro/mvputz

Zollinger-Ellison syndrome and Barrett's esophagus. By these means, proton-pump inhibitors (PPIs) became first-line pharmacological agents in treating the above-mentioned pathological conditions [1]. The therapeutic regimen that combines two antimicrobials (clarithromycin and amoxicillin or metronidazole) with a PPI is well-established in the eradication management of *Helicobacter pylori* infection [2].

In clinical practice PPIs with chemical structures of substituted benzimidazoles are currently used [3]. However, there are several PPIs that have been synthesized, some of them with other chemical structures: tenatoprazole (marked as TU-199) consists of one imidazopyridine ring connected to a pyridine ring by a sulfinylmethyl chain [2].

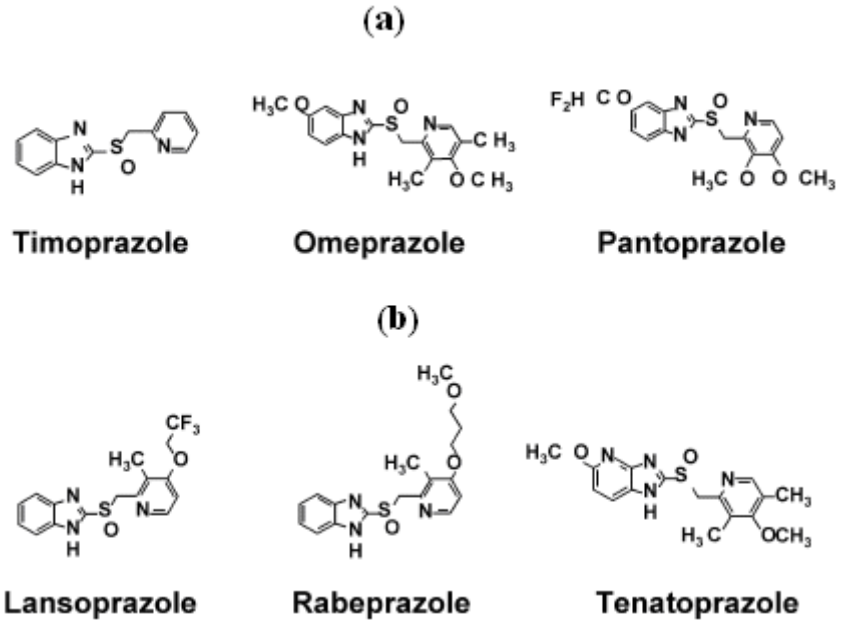

Figure 1: Proton pump types of inhibitors based on benzimidazole and imidazopyridine rinds in (a) and (b) respectively [1, 2].

Gastric parietal cells are responsible for acid secretion, controlled through several food-stimulated and neuroendocrine pathways mediated by gastrin, histamine, pituitary adenylate cyclase-activating peptide and acetylcholine [4]. PPIs target the final effector in the acid secretion pathway, namely the gastric H^+/K^+-ATPase [4]. In acidic environment, PPIs as prodrugs are converted to active sulfenamide form; then, the active drug forms a covalent bond to cysteine residues of active proton pumps producing an almost complete inhibition of gastric acid secretion which lasts until new proton pump molecules are synthesized [5].

There are irreversible and reversible gastric PPIs [6]. Irreversible gastric PPIs are structurally mainly characterized by the presence of the substituted pyridine ring, of the substituted benzimidazole ring, and of the methylsulfinyl linking group [6]. They are highly effective in gastric acid secretion inhibition leading in prolonged administration to extreme acid suppression and achlorohydria. Their most severe side-effects are hypergastrinemia, gastric polips and gastric carcinoma. In order to avoid these side-effects, several reversible

PPIs are under development, often referred as acid pump antagonists, but none is yet marketed [6].

In the effort to develop new effective drugs with gastric anti-secretor activity, thiazolidines show potential to inhibit the gastric H^+/K^+-ATPase [7].

Nevertheless, among many competitive-inhibitory pathways of H^+/K^+-ATPase proton pump, while the main chemical-physical process involved regards the PPIs diffusion through parietal cells in the stomach, the steps concerned in the signal transduction have not yet been elucidated. In this context, the present study employs of the recent proposed Spectral-SAR algorithm [8-13] to predict the most active molecular structure among a given series of PPI compounds by means of pre-designed activity paths.

2. Basics of the Spectral-SAR Method

Basically, the applied S-SAR algorithm can be summarized by means of the so called S-SAR operator defined as a collection of successive operations:

$$\hat{O}_{S-SAR}:\left\{\begin{array}{ll} Det\left(\left|Y^{Obs}\right\rangle,\left|Y^{P}\right\rangle,\left|X_0\right\rangle,\left|X_1\right\rangle,...,\left|X_M\right\rangle\right)=0, & \left\|\left|Y^{Obs/P}\right\rangle\right\|, r_{S-SAR}^{ALGEBRAIC}, \\ \delta\left[A_{(\|\bullet\rangle\|,r)},B_{(\|\bullet\rangle\|,r)}\right]=0,\langle\alpha,\beta,\gamma,...\rangle, & A,B:ENDPOINTS \\ & \alpha,\beta,\gamma,...:SPECTRAL\ \ PATHS \end{array}\right\} \quad (1)$$

for N-molecules with M-structural characteristics input data arranged within the vectors $|X_0\rangle=|1_1\ldots1_N\rangle$, $|X_i\rangle$, $i=1,\ldots M$, together with the measured (observed) endpoints $|Y^{Obs}\rangle$, see Table 1, while when computing the predicted endpoints $|Y^{P}\rangle$ the scalar product based spectral norms $\left\||Y^{Obs/P}\rangle\right\|$ are used to calculate the algebraic correlation factor $r_{S-SAR}^{ALGEBRAIC}$ as their ratio, until the first minimum paths across the spectral norms-correlation factors are attained between the specific endpoints, see bellow and ref. [10].

Table 1. The spectral (vectorial) - SAR generic table of descriptors.

Activity	***Structural predictor variables***					
$\|Y^{Obs}\rangle$	$\|X_0\rangle$	$\|X_2\rangle$	…	$\|X_k\rangle$	…	$\|X_M\rangle$
y_1	1	x_{11}	…	x_{1k}	…	x_{1M}
y_2	1	x_{21}	…	x_{2k}	…	x_{2M}
⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_N	1	x_{N1}	…	x_{Nk}	…	x_{NM}

Unfolding the above operator, the classical multivariate (QSAR) equation is here replaced by its predicted vectorial version,

$$\left|Y^{P}\right\rangle=b_0\left|X_0\right\rangle+b_1\left|X_1\right\rangle+...+b_k\left|X_k\right\rangle+...+b_M\left|X_M\right\rangle$$

$$= \left| Y^{Obs} \right\rangle - \left| e \right\rangle \,, \tag{2}$$

while the optimization procedure, corresponding to the null-vector for errors in (2),

$$\left| e \right\rangle = \left| 0_1 \, 0_2 \ldots 0_N \right\rangle, \tag{3}$$

provides the associated determinant [8, 10]

$$\begin{vmatrix} \left| Y^P \right\rangle & \omega_0 & \omega_1 & \cdots & \omega_k & \cdots & \omega_M \\ \left| X_0 \right\rangle & 1 & 0 & \cdots & 0 & \cdots & 0 \\ \left| X_1 \right\rangle & r_0^1 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \\ \left| X_k \right\rangle & r_0^k & r_1^k & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \\ \left| X_M \right\rangle & r_0^M & r_1^M & \cdots & r_k^M & \cdots & 1 \end{vmatrix} = 0 \,, \tag{4}$$

where the components are calculated with the scalar product recipes

$$r_i^k = \frac{\left\langle X_k \left| \Omega_i \right. \right\rangle}{\left\langle \Omega_i \left| \Omega_i \right. \right\rangle}, \; \omega_k = \frac{\left\langle \Omega_k \left| Y^{Obs} \right. \right\rangle}{\left\langle \Omega_k \left| \Omega_k \right. \right\rangle}, \; k = \overline{0, M} \tag{5}$$

by means of the Gram-Schmidt algorithm:

$$\left| \Omega_0 \right\rangle = \left| X_0 \right\rangle; \; \left| \Omega_k \right\rangle = \left| X_k \right\rangle - \sum_{i=0}^{k-1} r_i^k \left| \Omega_i \right\rangle . \tag{6}$$

The advantage of the Spectral-SAR method stands in the following points: it allows direct correlation equation without appealing to the computational methods of solving large systems with differential equations; it provides a transparent, algebraically direct regression method; it employs the orthogonality optimization framework for data correlation; treats the data as vectors thus providing an equation representative not only to individual molecular actions but to an entire set of chemicals action on certain biological system.

Beside these, the Spectral-SAR approach of correlation opens new perspectives of analytical analysis of chemical-biological interaction. Actually, since the vectorial frame of computation is undertaken the so called vectorial norm of a certain model (endpoint) either measured-observed or computed-predicted may be introduced by the consecrated *N*-components generalized definition of the norm-scalar product connection rule:

$$\left\| Y^{Obs/P} \right\rangle \| = \sqrt{\sum_{i=1}^{N} (y_i^2)^{Obs/P}} \quad . \tag{7}$$

Next, the vectorial norm gives the opportunity in advancing the so called algebraic correlation factor measuring the relative "intensity" or "amplitude" of the predicted to measured norm activities [10-13]:

$$r_{S-SAR}^{ALGEBRAIC} = \frac{\left\| Y^{P} \right\rangle \|}{\left\| Y^{Obs} \right\rangle \|} \quad . \tag{8}$$

However, worth noting that the algebraic correlation factor gives systematic higher values respecting the assumed statistical one, computed on statistical definition of the data dispersion through the standard expressions [14]

$$r_{QSAR}^{STATISTIC} = \sqrt{1 - \frac{SR}{SQ}}, SR = \sum_{i=1}^{N} \left[y_i^{Obs} - y_i^{P} \right]^2 , \; SQ = \sum_{i=1}^{N} \left[y_i^{Obs} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right]^2 , \tag{9}$$

for each of the studied cases so far [10-13]. Such behavior is susceptible to hold in general as it will be next exposed. In fact, we have to prove the proposition:

Proposition (algebraic S-SAR vs. statistic Q-SAR correlations)*: in any correlation analysis, considering the observed and predicted activity data as vectors* $\left| Y^{Obs} \right\rangle$ *and* $\left| Y^{P} \right\rangle$ *with the associate norms through the (7) scalar product type, respectively, the algebraic correlation factor (8) always exceeds the statistical QSAR correlation factor (9):*

$$r_{S-SAR}^{ALGEBRAIC} \geq r_{QSAR}^{STATISTIC} \quad . \tag{10}$$

Proof: by straight algebraic manipulation, condition (10) firstly rewrites as:

$$\frac{\sum_{i=1}^{N} \left(y_i^2 \right)^{P}}{\sum_{i=1}^{N} \left(y_i^2 \right)^{Obs}} \geq \frac{\sum_{i=1}^{N} \left[y_i^{P} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right] \left[2 y_i^{Obs} - y_i^{P} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right]}{\sum_{i=1}^{N} \left[y_i^{Obs} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right]^2} \; ; \tag{11}$$

Then, it can be conveniently arranged so that to separate the *predicted* and *measured* terms in left and right side of the inequality, respectively:

$$\frac{\sum_{i=1}^{N} \left(y_i^2 \right)^{P}}{\sum_{i=1}^{N} \left[y_i^{P} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right] \left[2 y_i^{Obs} - y_i^{P} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right]} \geq \frac{\sum_{i=1}^{N} \left(y_i^2 \right)^{Obs}}{\sum_{i=1}^{N} \left[y_i^{Obs} - N^{-1} \sum_{i=1}^{N} y_i^{Obs} \right]^2} \tag{12}$$

Now, both the nominator and denominator of the left-hand side of (12) are further handled so that to highlight the respective terms appeared in the right-hand side of (12):

$$\frac{\sum_{i=1}^{N}\left(y_i^2\right)^{Obs}-\left[\sum_{i=1}^{N}\left(y_i^2\right)^{Obs}-\sum_{i=1}^{N}\left(y_i^2\right)^{P}\right]}{\sum_{i=1}^{N}\left[y_i^{Obs}-N^{-1}\sum_{i=1}^{N}y_i^{Obs}\right]^2-\sum_{i=1}^{N}\left[y_i^{Obs}-y_i^{P}\right]^2}\geq\frac{\sum_{i=1}^{N}\left(y_i^2\right)^{Obs}}{\sum_{i=1}^{N}\left[y_i^{Obs}-N^{-1}\sum_{i=1}^{N}y_i^{Obs}\right]^2}\ ; \quad (13)$$

In this moment there appears that the difference between the right and left sides of (13) relays in the presence of the residuum of the sums of the measured and predicted activity squares in the nominator,

$$\begin{aligned} RS &= \sum_{i=1}^{N}\left(y_i^2\right)^{Obs}-\sum_{i=1}^{N}\left(y_i^2\right)^{P} \\ &= \left\|\left|Y^{Obs}\right\rangle\right\|^2-\left\|\left|Y^{P}\right\rangle\right\|^2 \end{aligned} \quad (14)$$

and of the sum of the residues in the denominator, see *SR* definition in (9):

$$SR=\left\|\left|Y^{Obs}\right\rangle-\left|Y^{P}\right\rangle\right\|^2, \quad (15)$$

emphasizing on the importance the residues concept plays in the (Q)SAR correlation theory; it is therefore clear that for perfect correlation, i.e. $y_i^{Obs}=y_i^{P}$, $\forall i=\overline{1,N}$, there will be no residuum type recorded in (14) and (15) and the relation (13) becomes equality. However, finally, the inequality (13) is true when the *SR* subtraction exceeds the *RS* subtraction in the denominator and nominator, respectively:

$$SR\geq RS\ ; \quad (16)$$

This can be easily achieved trough performing the chain relationships by combining (14) and (15) figures:

$$\begin{aligned} SR-RS &= \left\langle Y^{Obs}-Y^{P}\middle|Y^{Obs}-Y^{P}\right\rangle-\left\langle Y^{Obs}\middle|Y^{Obs}\right\rangle+\left\langle Y^{P}\middle|Y^{P}\right\rangle \\ &= 2\left(\left\langle Y^{P}\middle|Y^{P}\right\rangle-\left\langle Y^{Obs}\middle|Y^{P}\right\rangle\right) \\ &= 2\left[\left\langle Y^{P}\middle|Y^{P}\right\rangle-\left(\left\langle Y^{P}\right|+\left\langle e\right|\right)\left|Y^{P}\right\rangle\right] \\ &= -2\left\langle e\middle|Y^{P}\right\rangle \\ &\geq -2\left\|\left|e\right\rangle\right\|\left\|\left|Y^{P}\right\rangle\right\| \\ &= 0 \end{aligned} \quad (17)$$

with the help of scalar-product norm based (7) definition, the basic observed-predicted activity (2) constraint, the consecrated Cauchy-Schwartz inequality, and the optimized null vector of errors (3). Therefore, since the validity of (17) revealed the inequality (16) is fulfilled, and consequently (13) to (10) inequalities are as well demonstrated this way proofing the proposition. ■

As a note worth pointing out that the algebraic correlation factor (8) measures the ratio of vectorial "lengths" of the predicted to observed intensity of the chemical-biological interaction; moreover, this is sustained also by the final endpoint status carried by the observed activity vector; in fact any correlation analysis likes to find the proper structural combination factors and therefore also the mechanisms able to predict (or modeling) the observed endpoint as closely as possible. As such the vectorial picture assures the intuitive phenomenology in which the biological activity and its predicted models are represented by vectors and of their lengths (norms) with the observed activity being the highest over all models in the set of predicted mechanisms. This way the Spectral-SAR algorithm produces a correlation factor, the algebraic one (8), that provides both a better correlation and phenomenology respecting the statistical classical approach; this contribution however opens an epistemological discussion through the question: which correlation is the real one?

Here we only stipulate that the algebraic and statistical pictures are qualitatively different since they are based on vectorial length of predicted against the observed biological activities and on their dispersion, respectively.

In short, the Spectral-SAR analysis employs the idea of "amplitude" or "intensity" or "length" of chemical-biological interaction and bonding. In this context, appears also the idea of introducing *the least path principle* that selects the optimal (shortest) paths among all computationally tested models towards the measured endpoint:

$$\delta[A,B]=0;\quad A,B:ENDPOINTS\,. \tag{18}$$

Analytically, the paths [*A*, *B*] are computed via the Euclidian distances between each consecutive two points of a certain spectral path along the models in the abstract space of vectorial norms-correlation (recommended algebraically but being available also for statistically) factors:

$$[A,B]=\sqrt{\left(\left\|Y^{B}\right\rangle\right\|-\left\|Y^{A}\right\rangle\right\|\Big)^{2}+\left(r_{B}^{\substack{STATISTIC/\\ALGEBRAIC}}-r_{A}^{\substack{STATISTIC/\\ALGEBRAIC}}\right)^{2}}\quad . \tag{19}$$

The optimized spectral paths out of all set of possible "ergodic" combinations that connect the endpoints belonging to different modeling correlation classes, deliver the spectral hierarchies predicted; these are used to build a mechanistic picture of the molecular (ligand or effector) interaction with the substrate (organism or receptor) host.

The remaining issue is to use the Spectral-SAR methodology in order to predict the 3D configuration of the effector-receptor binding through the predicted endpoints, within norms - algebraic factors' space, employing minimum spectral path condition (18) targeting the minimum residuum of the predicted respecting observed chemical-biological activity.

The first attempt of this goal is in the next section tested on special case of H^+K^+-ATPase inhibitory activity.

3. Results for anti-ulcer compounds

Application of the above Spectral-SAR methodology is here exposed on a selected series of 4-Indolyl, 2-guanidinothiazole derivatives in the Table 2 with the general structure of Figure 2.

The data are arranged in a vectorial manner - see the Table 1, allowing the direct implementation of S-SAR algorithm. However, as structural parameters the standard Hansch correlation (20) is assumed through evaluating the hydrophobic (*LogP*), electronic (*POL*arizability) and steric (total energy E_{TOT}) contributions on predicted PPI activity.

$$A = b_0 + b_1\begin{pmatrix} hydrophobic \\ descriptor \end{pmatrix} + b_2\begin{pmatrix} electronic \\ descriptor \end{pmatrix} + b_3\begin{pmatrix} steric \\ descriptor \end{pmatrix} \tag{20}$$

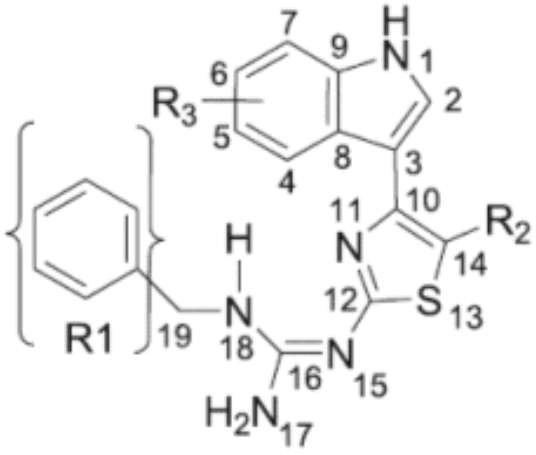

Figure 2: General scheme of 4-Indolyl, 2-guanidinothiazole derivatives.

Before going into particular analysis and results, let's quote that for observed activity we use the logarithm of the inverse of the IC50 measured concentration since it offers better discrimination (in terms of sign) for the low and high affinity for effector-receptor binding:

$$A = \log(1/IC_{50})\begin{cases} < 0 \ \ldots & IC_{50} > 1 \ \ldots \ low \ \ activity \\ > 0 \ \ldots & 0 < IC_{50} < 1 \ \ldots \ high \ \ activity \end{cases} \tag{21}$$

Table 2. Observed 4-Indolyl, 2-guanidinothiazole derivatives' inhibitory activities, along considered structural descriptors (hydrophobicity, polarizability, and total energy) computed with the help of HyperChem [15], in modeling the H^+K^+-ATPase gastric protonic pump molecular mechanisms [7].

No	*Compounds*				$\|Y^{Obs}\rangle$=*A*=	$\|X_1\rangle$=	$\|X_2\rangle$=	$\|X_3\rangle$=E_{TOT}
	Name	*R1*	*R2*	*R3*	*Log(1/IC50)*	*LogP*	*POL[Å^3]*	*[kcal/mol]*
1	4-indolyl-2-methylguanidinothiazole	H	H	H	-2.1972	-1.0	30.27	-14.9839
2	4-indolyl-2-methylheptylguanidinothiazole	C_7H_7	H	H	-0.4055	-0.51	41.76	-29.4849
3	5-methyl-4-indolyl-2-methylguanidinothiazole	H	Me	H	-2.0281	-0.98	32.10	-27.3296
4	4-(5-methoxyindolyl)-2-methylguanidinothiazole	H	H	5-OMe	-2.9444	-2.00	32.74	-19.6745
5	4-(5-methoxyindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-OMe	-1.4586	-0.48	44.24	-35.2272
6	4-(5-benzyloxyindolyl)-2-methylguanidinothiazole	H	H	5-OC_7H_7	-0.1823	-0.51	42.40	-34.1524
7	4-(5-benzyloxyindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-OC_7H_7	-0.5878	1.01	53.90	-44.7714
8	4-(2-methylindolyl)-2-methylguanidinothiazole	H	H	2-Me	-2.1972	-0.98	32.10	-21.9728
9	4-(2-methylindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	2-Me	***-0.5306***	0.54	43.60	-30.9985
10	5-methyl-4-(2-methyl, 5-cloroindolyl)-2-methylguanidinothiazole	H	Me	2-Me, 5-Cl	0.5108	-1.17	35.87	-31.2836
11	5-methyl-4-(2-methyl, 5-cloroindolyl)-2-methylheptylguanidinothiazole	C_7H_7	Me	2-Me, 5-Cl	-0.0953	0.34	47.36	-39.1314
12	4-(4-methylindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	4-Me	-1.1939	0.66	43.60	-38.6181
13	4-(5-methylindolyl)-2-methylguanidinothiazole	H	H	5-Me	-0.5878	-0.85	32.10	-20.2655
14	4-(5-methylindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-Me	***0.0619***	0.66	43.60	-29.0775
15	4-(6-methylindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	6-Me	***-0.4700***	0.66	43.60	-28.6606

Table 2. (Continued)

16	4-(7-methylindolyl)-2-methylguanidinothiazole	H	H	7-Me	-2.0281	-0.85	32.10	-20.8977
17	4-(7-methylindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	7-Me	***0.0***	0.64	41.76	-27.5365
18	4-(5-cloroindolyl)-2-methylguanidinothiazole	H	H	5-Cl	-0.5878	-1.23	32.20	-21.3425
19	4-(5-cloroindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-Cl	0.3567	0.29	43.69	-29.6063
20	5-methyl-4-(5-cloroindolyl)-2-methylguanidinothiazole	H	Me	5-Cl	**0.5276: *max***	-1.20	34.03	-28.0026
21	5-methyl-4-(5-cloroindolyl)-2-methylheptylguanidinothiazole	C_7H_7	Me	5-Cl	0.3567	0.32	45.53	-37.9271
22	4-(5-bromoindolyl)-2-methylguanidinothiazole	H	H	5-Br	0.0408	-0.95	32.89	-21.1396
23	4-(5-bromoindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-Br	0.3567	0.56	44.39	-28.7985
24	4-(5-fluoroindolyl)-2-methylguanidinothiazole	H	H	5-F	-2.0015	-1.61	30.18	-20.5329
25	4-(5-fluoroindolyl)-2-methylheptylguanidinothiazole	C_7H_7	H	5-F	***-0.9933***	-0.09	41.67	-30.8076
26	4-(5-methylcetylindolyl)-2-methylguanidinothiazole	H	H	5-CO_2Me	***-1.9021***	-1.59	34.66	-27.4917
27	4-(5-carboximethylindolyl)-2-ethylbenzeneguanidinothiazole	C_7H_7	H	5-CO_2Me	-1.1314	-0.07	46.16	-37.7917
28	4-(5-amidoethylindolyl)-2-ethylbenzeneguanidinothiazole	C_7H_7	H	5-NHCOEt	**-3.1697: *min***	-0.75	48.71	-33.8852

This rule further opens the way in deciding at once which molecule of Table 2 is the most active one; it follows that the molecule no. 20 (5-methyl-4-(5-cloroindolyl)-2-methylguanidinothiazole) displays the most positive record throughout the entire considered series. However, in molecular design we are searching for molecules that optimally act through controlled or predicted mechanisms. In this respect, we like to test whether the Hansch assumed mechanisms (seen as a combination of transduction by *LogP*, electrostatic interaction by *POL*, and fine tuning of covalent bindings by E_{TOT}) eventually leads with the same results or predict something different. Such endeavor can be fully accomplished by applying the S-SAR algorithm by means of the minimum spectral paths analysis; it produces the results presented in the Table 3 when the S-SAR recipe was respective engaged for various endpoint models for single-, doubled- and all- structural included effects in (20).

Table 3. Spectral-SAR multi-linear equations for all possible correlation models considered from data of Table 2; $||\ Y^{Obs}\ ||$ = 7.26044 is computed from Table 2 with the rule (7), while $r_{Algebraic}$ and $r_{Statistic}$ are evaluated upon eqs. (8) and (9), respectively.

Model	Variables	S-SAR Equation	$\|\|\|Y^P\rangle\|\|$	$r_{Algebraic}$	$r_{Statistic}$
Ia	$\|X_0\rangle$, $\|X_1\rangle$	$\|Y\rangle^{Ia} = -0.617953\|X_0\rangle + 0.644409\|X_1\rangle$	5.42199	0.746785	0.505266
Ib	$\|X_0\rangle$, $\|X_2\rangle$	$\|Y\rangle^{Ib} = -2.84242\|X_0\rangle + 0.0497704\|X_2\rangle$	4.9347	0.67967	0.306765
Ic	$\|X_0\rangle$, $\|X_3\rangle$	$\|Y\rangle^{Ic} = -2.26235\|X_0\rangle - 0.0478985\|X_3\rangle$	4.95654	0.682677	0.317809
IIa	$\|X_0\rangle$, $\|X_1\rangle, \|X_2\rangle$	$\|Y\rangle^{IIa} = 1.2815\|X_0\rangle + 0.929486\|X_1\rangle - 0.0451665\|X_2\rangle$	5.50092	0.757657	0.531819
IIb	$\|X_0\rangle$, $\|X_1\rangle, \|X_3\rangle$	$\|Y\rangle^{IIb} = -0.55429\|X_0\rangle + 0.655559\|X_1\rangle + 0.00204383\|X_3\rangle$	5.4223	0.746828	0.505372
IIc	$\|X_0\rangle$, $\|X_2\rangle, \|X_3\rangle$	$\|Y\rangle^{IIc} = -2.56541\|X_0\rangle + 0.0192239\|X_2\rangle - 0.0321241\|X_3\rangle$	4.96627	0.684018	0.322626
III	$\|X_0\rangle$, $\|X_1\rangle, \|X_2\rangle$, $\|X_3\rangle$	$\|Y\rangle^{III} = 2.33209\|X_0\rangle + 1.03407\|X_1\rangle - 0.120533\|X_2\rangle - 0.0680251\|X_3\rangle$	5.62022	0.774088	0.570269

Worth noted that while the statistical correlation factors (9) for computed models in Table 3 give quite modest values, indicating practically no correlation effects, the algebraic frame, specifically to S-SAR analysis, deliver systematically higher values, and consequently improved correlations, in each endpoint predicted case. This is a typical behavior within the vectorial approach, already anticipated by above S-SAR correlation proposition.

Next, having the vectorial norms and algebraic (or statistic) correlation factors one can unfold the spectral-path calculations, based on iterative application of the eulerian formula (19). The hierarchy of alpha, beta and gamma paths is established on the ground of the minimum path principle, both at global and local levels: firstly, the shortest path is selected; if two paths are equal the least will be set along which the distance between the endpoint norms of the first two models along the equivalent paths is minimum (this is based on heuristic, however natural, principle of "first minimum movement" as being favorite); if, by chance, also these distances are equal the chosen path will be that starting on the model with the closest predicted norm to the (experimentally) observed one.

With the present least global-to-local selection algorithm the results processed for data of Table 3 are Table 4 displayed. Note that, although no relevant for the S-SAR study the statistical correlation factors provide the same spectral path hierarchy as those algebraically based.

Table 4. Synopsis of paths connecting the S-SAR models of Table 3 in the norm-correlation spectral-space with the α, β, and γ first least spectral paths highlighted.

Path	Value	
	Algebraic	**Statistic**
Ia-IIa-III	0.200101	0.208619
Ia-IIb-III	**α** 0.200101	**α** 0.208615
Ia-IIc-III	1.12014	1.19022
Ib-IIa-III	**γ** 0.691988	**γ** 0.734646
Ib-IIb-III	0.691988	0.734596
Ib-IIc-III	0.691988	0.734781
Ic-IIa-III	0.669948	0.710281
Ic-IIb-III	0.669948	0.710398
Ic-IIc-III	**β** 0.669948	**β** 0.710129

However, this approach has to be further completed with a mechanism of identifying the most active molecule trough the selected spectral paths. This can be reached by considering two more stages in selecting the molecules along the spectral paths. Actually, in the first stage the residues of all predicted models' activities (computed upon the S-SAR equations of Table 3) against the observed activity of molecules of Table 2 are evaluated, and the Table 5 produced. Nevertheless this is due the huge role the residues acquire in Spectral-SAR correlation - see the left side of (13) with the residues types (14) and (15) around the occasion of algebraic versus statistical previous discussion.

Table 5. Residual activities $A_i - Y_i^P$ of the compounds of Table 2 for the Spectral-SAR models of Table 3 ordered according with the alpha, beta and gamma paths of Table 4.

No.	*Models*						
	α		*β*		*γ*		
	Ia	***IIb***	***Ic***	***IIc***	***Ib***	***IIa***	***III***
1	-0.934838	-0.956726	-0.652554	-0.695041	-0.861333	-1.18202	-0.865965
2	0.541102	0.543387	0.444569	0.409946	0.358506	0.673195	0.817532
3	-0.778626	-0.775505	-1.0748	-0.957714	-0.783313	-0.948854	-1.33679
4	-1.03763	-1.03878	-1.62443	-1.6404	-1.73147	-0.888172	-0.600457
5	-0.531331	-0.517643	-0.883578	-0.875296	-0.818025	-0.295776	-0.358289
6	0.764302	0.776127	0.444203	0.470903	0.549852	0.925302	0.800366
7	-0.6207	-0.604119	-0.469931	-0.496797	-0.428007	-0.373602	-0.51315
8	-0.947726	-0.955553	-0.987312	-0.954732	-0.952413	-1.11795	-1.14149
9	***XIV:-0.260628***	-0.266956	**XII:0.24697**	**VI:0.200851**	**III:0.141828**	-0.344759	-0.274524
10	1.88271	1.89603	1.27471	1.38169	1.56795	1.93693	1.58402
11	0.303554	0.316078	0.292716	0.302607	0.389992	0.446265	0.267553
12	-1.00126	-0.99335	-0.781297	-0.707222	-0.521472	-1.1196	-1.58024
13	0.577901	0.565135	0.703865	0.709513	0.656987	0.370612	0.449618
14	0.254543	**XI:0.242951**	0.931483	0.855061	0.734328	0.136203	0.324564
15	-0.277357	-0.289801	0.419552	0.336553	0.202428	-0.395697	***V:-0.178976***
16	-0.862399	-0.873873	-0.766717	-0.751096	-0.783313	-1.06969	-1.03369
17	**VII:0.205531**	0.191012	0.943395	0.878036	0.764006	**I:0.00978637**	**IV:0.166392**
18	0.822776	0.816448	0.652278	0.672993	0.65201	0.728334	0.781355
19	0.787774	0.781388	1.20095	1.13114	1.02465	0.778978	0.976846
20	*1.91884*	*1.92579*	*1.44867*	*1.53926*	*1.67633*	*1.8985*	*1.63325*
21	0.768442	0.778728	0.802401	0.828474	0.933071	0.8342	0.601582
22	1.27094	1.26108	1.2906	1.29485	1.24627	1.12784	1.21739
23	0.613784	0.602736	1.23965	1.14364	0.989809	0.559633	0.836971
24	-0.346048	-0.349794	-0.722643	-0.675867	-0.661154	-0.423398	-0.427802
25	-0.31735	-0.317044	***VIII:-0.206586***	***IX:-0.218615***	***X:-0.224815***	-0.309054	-0.305402
26	0.259537	***XIII:-0.249283***	-0.956559	-0.886135	-0.784725	***II:-0.140242***	-0.282469
27	-0.468338	-0.453981	-0.679214	-0.667388	-0.586384	-0.262946	-0.398085
28	-2.06844	-2.05448	-2.5304	-2.62922	-2.7516	-1.55402	-1.16012

In the second stage, from the residues Table 5 there are selected the least positive (maximum) and negative (minimum) values for each endpoint along the optimized spectral paths (α, β, and γ) because these will clearly indicate the best predicted-measured fit. Now, the obtained series of maximum & minimum residue values across α, β, and γ spectral paths is ordered from the smallest ahead and for each of it the respective observed activity is identified. The molecules with null and negative observed activities are considered unfavorable (or hard favorable) in binding. The search continues until the first molecule of the ordered set is found with the closest to the higher observed activity in Table 2.

In the present analysis, it comes out that the molecule no. 14 of Table 2, namely the 4-(5-methylindolyl)-2- methylheptylguanidinothiazole, provides such behavior. Note that if, by chance, there are two candidate molecules furnishing the same observed activity will be preferred that one belonging to the spectral pathway with superior degree (α>β>γ). If still, also by this constraint the molecules belongs to the same path (say they both display equal positive and negative residual values respecting the same observed activity) then the common

skeleton resulting from their superposition will be considered as the optimum-by-spectral-paths active molecule.

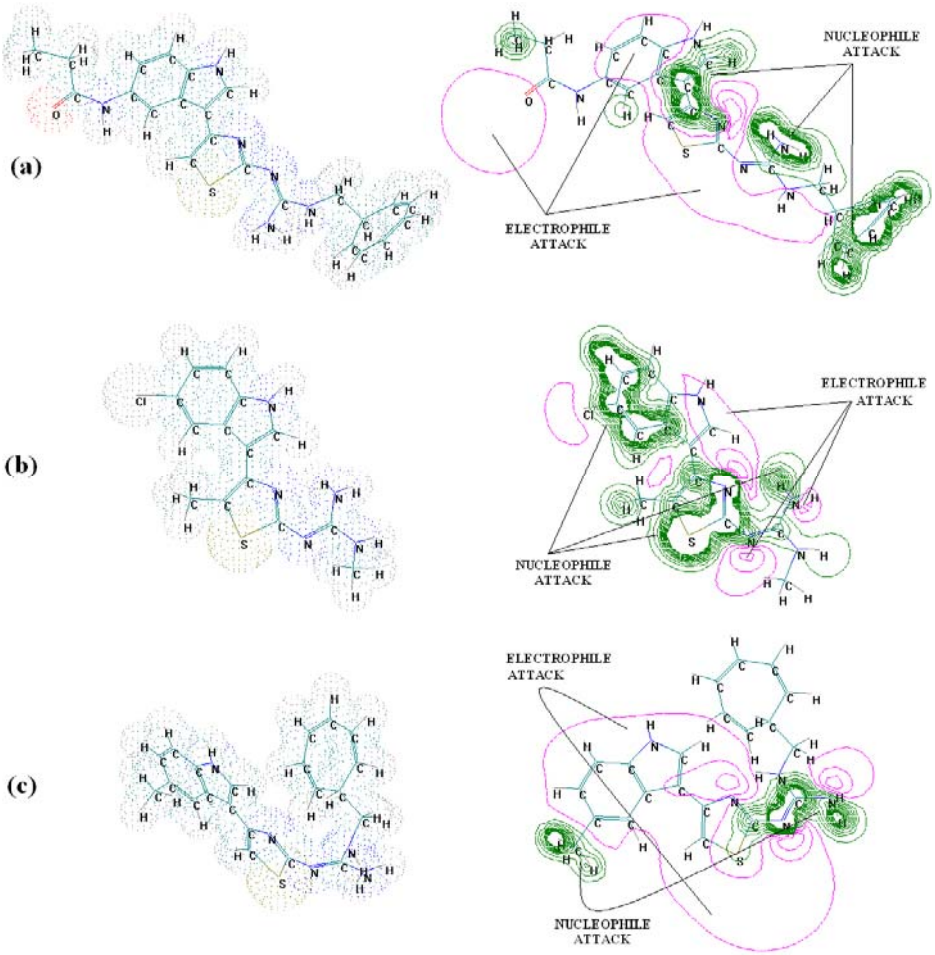

Figure 3: The electronic basins (in left sides) and the associated electrostatic potential contours (in right sides) for (a) molecule no. 28 of Table 2, emphasizing minimum action and maximum misfit; (b) molecule no. 20 of Table 2, with maximum observed action and absolute minimum misfit; and (c) molecule no. 14 of Table 2, displaying optimized maximum activity by predicted mechanism action.

In fact, this kind of completing of S-SAR algorithm leaves with the predicted 3D molecular structure that throughout a certain structural mechanism models the optimum attack on envisaged biological sites. It may be appropriately called as 3D-S-SAR method.

However, in the present case, there is clear that the resulted 3D-S-SAR molecule is significantly different by that simply picked up from Table 2 grounded only on the highest observed activity. In this regard, since the molecule of Table 2 with the smallest activity value, the molecule no. 28, 4-(5-amidoethylindolyl)-2- ethylbenzeneguanidinothiazole, still produces an observable activity, it means that the actual founded molecule, no. 14 in Table 2, assure about 87.4 %, of the maximum possible activity in the series.

Moreover, the molecule selected via the computational 3D-S-SAR scheme is expected to better satisfy the controlling and designing needs of the drugs along a specific pathway interactions.

In other words, the molecular structure that produces an appreciable biological activity through the alpha (spectral) path of chemical-biological interaction, i.e. through the molecular dispersion (*LogP*), followed by steric-covalent tuning of the ligand-receptor (E_{TOT}) approaching and, in the final, by electrostatic filed interaction (*POL*) here as the electrophile attack - is that identified as molecule no. 14 out of series in Table 2.

This is not surprising at all since the representations in the Figure 3. At close inspection to the electrostatic contours for the molecules with the minimum, highest, and the 3D-S-SAR provided activities, nos. 28, 20, and 14 in Table 2, respectively, there is clear that the last one is characterized by the largest area of electrophile attack over the basic benzothiazolyl ring of the three analyzed compounds. With this feature there seems that the 4-(5-methylindolyl)-2-methylheptylguanidinothiazole molecule better responds to the inner mechanisms of protonation upon entering the lower pH gradient that is present near the secretory canaliculus. The molecules with minimum and maximum observed activities indeed show somewhat complementary electrophile/nucleophile areas in Figure 3, (a) and (b), respectively, being perhaps involved in many indirect or complex signaling pathways.

Further studies may focus on finding the appropriate structural factors that being involved in the 3D-Spectral-SAR correlation and path analysis to predict the molecular mechanism recovering the maximum observed activity at the best.

6. Conclusion

In quantitative structure-activity relationship studies (QSARs), beside searching for the best correlation of predicted with the observed data, stands also the aim of identifying the 3-dimensional molecular structure configuration that fits at the best with the envisaged receptor throughout a given (thus susceptible of being controlled) mechanism of action. This way the molecular design for a specific chemical-biological interaction pathway can be used as the benchmark for the new synthesized chemicals and drugs.

However, unfortunately, until now there was not approached the 3D predicted structure from the correlation itself but only as a prerequisite for further correlation through validations and tests. To balance this miss, this work opens the way of extracting the optimized 3D molecular ligand structure locking the targeted receptor's activity fit as a result of QSAR performed correlations on a given series of analogs.

To this end the Spectral-SAR algorithm was here tested in delivering such 3D structural information. It starts with the consecrated vectorial analysis [10-13] until finding the main least spectral paths connecting the considered endpoint computed models for a paradigmatic correlation analysis, here of Hansch type. Then, for each proposed model, the residues of the predicted activities against the observed ones were evaluated for each molecule of the working series to build a restricted minimum-to-maximum residue series for candidate compounds. The final cut extracts that molecule from the residues ordered series that displays the closest (positive) activity to the maximum observed in the original (complete) series of compounds. This so called 3D-Spectral-SAR scheme allows further identification of molecular structure that is intrinsically controlled by the envisaged structural factors (here

LogP, *POL*, and E_{TOT}) and by the associated pathways towards the receptor or substrate target. However, this by-spectral-SAR predicted search of molecular structure may often be different by the molecule in original series displaying the maximum observed activity.

This was the case also for the actual study of the inhibitory activity of 4-Indolyl, 2-guanidinothiazole derivatives upon the H^+K^+-ATPase proton pump in parietal cells of the stomach. Actually, there was found that the molecule selected by the 3D-Spectral-SAR method shows a large area of electrophilic interaction which eventually links with the protonation role the optimal H^+K^+-ATPase inhibitory activity has to have; instead, the molecule with the highest observed activity in the input analogs series merely poses an equilibrium among the electrophilic and nucleophilic active sites suggesting that it is involved in a more complex physico-chemical signaling activity molecular or paths mechanisms in the cell than possible to analytically identify by the present analysis.

Nevertheless, at this point the temporal issue of the chemical-biological interaction may be assessed in order to decide whether the 3D Spectral-SAR molecular selecting algorithm is associated also with the shortest temporal range of action. Such studies may be unfolded by involving the recent logistical version of the enzyme-substrate kinetics [13, 16-18] and are currently in progress.

Acknowledgement

The financial support of the Romanian National Council of Scientific Research in Universities – CNCSIS is kindly thanked.

References

[1] Horn J., The proton-pump inhibitors: similarities and differences. *Clin. Ther.* **2000**, 22, 266-280.

[2] Scarpignato C., New drugs to suppress acid secretion: current and future developments. *Drug Discov. Today: Ther. Strategies* **2007**; doi:10.1016/j.ddstr.2007.09.003.

[3] Vanderhoff B.T., Tahboub R.M., Proton pump inhibitors: an update. *Am. Fam. Physician.* **2002**, 66, 273-280.

[4] Sachs G., Shin J.M., Howden C.W., Review article: the clinical pharmacology of proton pump inhibitors. *Aliment. Pharmacol. Ther.* **2006**, 23(Suppl. 2), 2-8.

[5] Marchetti F., Gerarduzzi T., Ventura A., Proton pump inhibitors in children: a review. *Digestive and Liver Disease* **2003**, 35:738-746.

[6] Jain K.S., Shah A.K., Bariwal J., Shelke S.M., Kale A.P., Jagtap J.R., Bhosale A.V., Recent advances in proton pump inhibitors and management of acid-peptic disorders. *Bioorg. Med. Chem.* **2007**, 15, 1181-1205.

[7] Prabhakar Y.S., Solomon V.R., Gupta M.K., Katti S.B., QSAR studies on thiazolidines: a biologically privileged scaffold. *Top. Heterocycl. Chem.* **2006**, 4, 161-249.

[8] Putz M.V., A Spectral Approach of the Molecular Structure – Biological Activity Relationship Part I. The General Algorithm, *Ann. West Univ. Timişoara, Series Chem.*, **2006**, 15, 159-166.

[9] Putz M.V., Lacrămă A.M., A Spectral Approach of the Molecular Structure – Biological Activity Relationship Part II. The Enzymatic Activity, *Ann. West Univ. Timişoara, Series Chem.*, **2006**, 15, 167-176.

[10] Putz M.V., Lacrămă A.M., Introducing Spectral Structure Activity Relationship (S-SAR) Analysis. Application to Ecotoxicology, *Int. J. Mol. Sci.*, **2007**, 8, 363-391.

[11] Lacrămă A.M., Putz M.V., Ostafe V., A Spectral-SAR Model for the Anionic-Cationic Interaction in Ionic Liquids: Application to *Vibrio fischeri* Ecotoxicity, *Int. J. Mol. Sci.*, **2007**, 8, 842-863.

[12] Putz M.V., Lacrămă A.-M., Ostafe V., Spectral-SAR Ecotoxicology of Ionic Liquids. The *Daphnia magna* Case, *Res. Lett. Ecol.*, **2007**, Article ID12813/5 pages, doi: 10.1155/2007/12813 (*http://www.hindawi.com/journals/rleco/q4.2007.html*).

[13] Lacrămă A.M., Putz M.V., Ostafe V., Designing a Spectral Structure-Activity Ecotoxico-Logistical Battery, in "*Advances in Quantum Chemical Bonding Structures*", Mihai V. Putz (Ed.), Transworld Network Research, Kerala, India, **2008**, in press (*http://www.trnres.com/putz.htm*).

[14] Miller J.N., Miller J.C., Statistics and Chemometrics for Analytical Chemistry, Fourth Edition, Pretice Hall, New York, **2000**.

[15] Hypercube, Inc. (**2002**) HyperChem 7.01 [Program package].

[16] Putz M.V., Lacrămă A.-M., Ostafe V. Full Analytic Progress Curves of the Enzymic Reactions in Vitro, *Int. J. Mol. Sci.*, **2006**, 7, 469-484.

[17] Putz M.V., Lacrămă A.-M. Enzymatic control of the bio-inspired nanomaterials at the spectroscopic level, *J. Optoel. Adv. Mat.*, **2007**, 9(8), 2529 – 2534.

[18] Putz M.V., Lacrămă A.-M. Ostafe V. Introducing logistic enzyme kinetics, *J. Optoel. Adv. Mat.*, **2007**, 9(9), 2910 – 2916.