STUDIA UNIVERSITATIS BABEŞ-BOLYAI, CHEMIA, LIII, 2, 2008

SPECTRAL-SAR: OLD WINE IN NEW BOTTLE

MIHAI V. PUTZ[a,*] and ANA-MARIA PUTZ (b. LACRĂMĂ)[b]

ABSTRACT. With the aim providing an efficient and simple scheme of correlation between the chemical structures – on one hand – and the chemical properties or biological activities – on other hand – the available QSAR schemes are firstly grouped within classical, 3-dimensional, decisional, and orthogonal schemes while shortly reviewed their characteristics. Then, particular attention is paid on the recent orthogonal QSAR method, namely Spectral-SAR (S-SAR) algorithm, of which mathematical principles along its computational advantages are exposed. Application of S-SAR methodology is also given by modeling the mechanistic hierarchy of some xenobiotics ecotoxicology on the battery of *Protista*, *Monera*, and *Animalia* species kingdoms.

MOTIVATION

One of the main challenges in actual conceptual and computational chemistry stands the (quantum) transfer of the chemical interaction from local to global interaction in complex (biomolecular) systems. Although many attempts were advanced, based on reactivity and topological indices, the correlation analysis still remains to pay tribute to traditional QSAR (quantitative structure-activity relationship) statistical schemes. However, there is widely believed that while the chemical information is somehow iteratively carried from atoms to molecules to system a similar mathematical algorithm should follow the chemical bond and bonding from an intrinsic (structural or causal) to a manifested (observed or effector-receptor measured) levels [1].

Nevertheless, the superposition principle in quantum chemistry is here employed to the liner combination of intrinsic states assume to be orthogonal, i.e. carrying independent non-redundant information. In fact, the data from molecular real space (inherently with a degree of inter-correlation) has to be (at least) formally orthogonalized in an abstract space, see Figure 1, for a better approaching of bonding (correlation) analysis that has to be finally remapped in the real space however with the solved causality of actions. The present work reviews such an attempt.

---

[a] *Laboratory of Computational and Structural Physical Chemistry, Chemistry Department, West University of Timişoara, Pestalozzi Str. No.16, Timişoara, RO-300115, Romania; Tel: +40/256/592633, Fax: +40/256/592620*

[*] *E-mails: mv_putz@yahoo.com or mvputz@cbg.uvt.ro, Web: http://www.cbg.uvt.ro/mvputz*

[b] *Institute of Chemistry Timişoara of the Romanian Academy, 24 Mihai Viteazul Avenue, RO-300223, Timişoara, Romania*

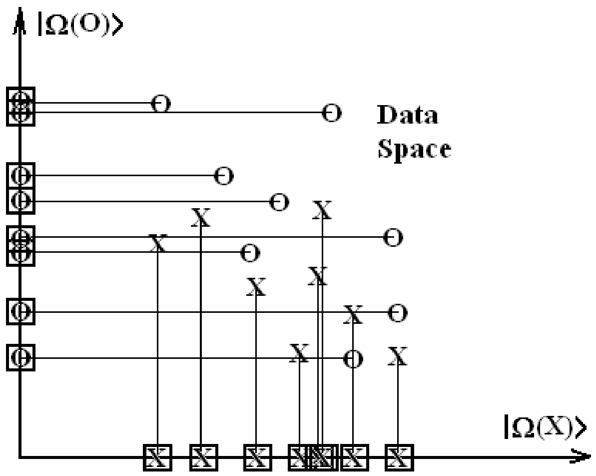

Figure 1. Generic mapping of data space containing the vectorial sets $\{|X\rangle, |O\rangle\}$ into orthogonal basis $\{|\Omega(X)\rangle, |\Omega(O)\rangle\}$ [1].

S-SAR METHODOLOGY

In that way, using the ideal condition of orthogonality of a QSAR correctional analysis the data space is considered as a vectorial one where the columns are treated like vectors, see Table I.

Table 1.
Spectral (vectorial) description of conventional SAR data [1-3].

Activity	*Structural predictor variables*					
$\|Y\rangle$	$\|X_0\rangle$	$\|X_1\rangle$	$\cdots$	$\|X_k\rangle$	$\cdots$	$\|X_M\rangle$
y_1	1	x_{11}	$\cdots$	x_{1k}	$\cdots$	x_{1M}
y_2	1	x_{21}	$\cdots$	x_{2k}	$\cdots$	x_{2M}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
y_N	1	x_{N1}	$\cdots$	x_{Nk}	$\cdots$	x_{NM}

Than, the vectorial space of data $\{|X_0\rangle, |X_1\rangle, ..., |X_k\rangle, ..., |X_M\rangle\}$ is transformed in an orthogonal one $\{|\Omega_0\rangle, |\Omega_1\rangle, ..., |\Omega_k\rangle, ..., |\Omega_M\rangle\}$ with the aid of iterative Gram-Schmidt procedure;

$$\left|\Omega_k\right\rangle = \left|X_k\right\rangle - \sum_{i=0}^{k-1} r_i^k \left|\Omega_i\right\rangle, \quad r_i^k = \frac{\left\langle X_k \middle| \Omega_i\right\rangle}{\left\langle \Omega_i \middle| \Omega_i\right\rangle}$$

with the help of scalar product rule for arbitrary N-dimensional vectors

$\left|V\right\rangle = \left|v_1 ... v_N\right\rangle$ and $\left|W\right\rangle = \left|w_1 ... w_N\right\rangle$:

$$\left\langle V \middle| W\right\rangle = \sum_{i=1}^{N} v_i w_i$$

Now the correlation equation in orthogonal space takes the form

$$\left|Y\right\rangle = \omega_0\left|\Omega_0\right\rangle + \omega_1\left|\Omega_1\right\rangle + ... + \omega_k\left|\Omega_k\right\rangle + ... + \omega_M\left|\Omega_M\right\rangle, \quad \omega_k = \frac{\left\langle \Omega_k \middle| Y\right\rangle}{\left\langle \Omega_k \middle| \Omega_k\right\rangle},$$

$$k = \overline{0, M}.$$

Since the come back in the real space is realized through the specific coordinate transformation

$$\begin{cases}
\left|Y\right\rangle = \omega_0\left|\Omega_0\right\rangle + \omega_1\left|\Omega_1\right\rangle + ... + \omega_k\left|\Omega_k\right\rangle + ... + \omega_M\left|\Omega_M\right\rangle \\
\left|X_0\right\rangle = 1\cdot\left|\Omega_0\right\rangle + 0\cdot\left|\Omega_1\right\rangle + ... + 0\cdot\left|\Omega_k\right\rangle + ... + 0\cdot\left|\Omega_M\right\rangle \\
\left|X_1\right\rangle = r_0^1\left|\Omega_0\right\rangle + 1\cdot\left|\Omega_1\right\rangle + ... + 0\cdot\left|\Omega_k\right\rangle + ... + 0\cdot\left|\Omega_M\right\rangle \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\left|X_k\right\rangle = r_0^k\left|\Omega_0\right\rangle + r_1^k\left|\Omega_1\right\rangle + ... + 1\cdot\left|\Omega_k\right\rangle + ... + 0\cdot\left|\Omega_M\right\rangle \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\left|X_M\right\rangle = r_0^M\left|\Omega_0\right\rangle + r_1^M\left|\Omega_1\right\rangle + ... + r_k^M\left|\Omega_k\right\rangle + ... + 1\cdot\left|\Omega_M\right\rangle
\end{cases}$$

that will generate the searched correlation equation, which results in the actual S-SAR equation [1-3]:

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

thus providing the real activity-structural equation:

$$\left|Y\right\rangle = b_0\left|X_0\right\rangle + b_1\left|X_1\right\rangle + \ldots + b_k\left|X_k\right\rangle + \ldots + b_M\left|X_M\right\rangle$$

with the coefficients b_j, j=0, …, M, determined by expansion of above determinant along its first column.

Such algebraically approach of structure-activity correlation, allows the use of algebraically properties of the norm, here introduced as spectral norm,

$$\left\| \left|Y\right\rangle^{MEASURED/PREDICTED} \right\| = \sqrt{\sum_{i=1}^{N} (y_i^2)^{MEASURED/PREDICTED}}$$

together with new introduced correlation factor of algebraically type,

$$r_{S-SAR}^{ALGEBRAIC} = \frac{\left\| \left|Y\right\rangle^{P(REDICTED)} \right\|}{\left\| \left|Y\right\rangle^{M(EASURED)} \right\|}$$

Notably, the superiority of Spectral-SAR over traditional QSAR analysis was recently proved through the algebraic S-SAR correlation proposition [4]: in any vectorial analysis of Spectral-SAR type, considering the measuring and observed activity data as vectors $\left|Y^M\right\rangle$ and $\left|Y^P\right\rangle$ with the associate, while assuming the norm order

$$\left\| \left|Y^P\right\rangle \right\| \le \left\| \left|Y^M\right\rangle \right\|$$

the algebraic correlation factor will always exceed the standard QSAR statistical correlation factor , i.e. $r_{S-SAR}^{ALGEBRAIC} \ge r_{QSAR}^{STATISTIC}$, with

$$r_{QSAR}^{STATISTIC}=\sqrt{1-\frac{SR}{SQ}}\ ,\quad SR=\sum_{i=1}^{N}\left[y_i^M-y_i^P\right]^2\ ,$$
$$SQ=\sum_{i=1}^{N}\left[y_i^M-N^{-1}\sum_{i=1}^{N}y_i^M\right]^2\ .$$

The study recommended the use of algebraically definition instead of the statistical one taking in consideration that the first has the physically meaning of “length of action” while replacing the standard dispersion definition. While the dispersion is a consequence of quality of fitting the vectorial norm of “length of action” simple describe the degree with which a certain model approaches the measured or manifested chemical-biological interaction.

S-SAR Ecotoxicology Applications

The Spectral-SAR methodology allows considering the spectral pathway of computed endpoints (activities) [1, 5]:

$$[A,B]=\sqrt{\left(\left\|Y^B\right\rangle\right\|-\left\|Y^A\right\rangle\right\|\right)^2+\left(r_B^{STATISTIC/ALGEBRAIC}-r_A^{STATISTIC/ALGEBRAIC}\right)^2}$$

Moreover, the introduction of spectral way allows the formulation for the first time of a principle of optimization of pathways between the correlation models,

$$\delta[A,B]=0;\quad A,B:ENDPOINTS$$

reflecting the optimisation of pathways between different endpoints of an open chemical-biological system. This principle says that from the all ways unifying the endpoints of a chemical-biologically system that pathway having the minimal length while finding the maximum possible combinations of the considered structural parameters is selected.

As such, the spectral scheme of the molecular mechanisms is obtained as a hierarchy of action influences of a series of chemicals on a certain species. Actually, the present Spectral-SAR model was tested for Hansch type regression by correlating the lipophylicity, the electronic polarizability and the total energy at the optimal configuration

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

with the ecotoxicity from different molecular levels and in various species.

At the unicellular level the ecotoxicity of the Tetrahymena pyriformis species was modeled employing the minimal spectral path procedure, i.e. ordering all the possible paths which connect endpoints from each mono-, bi- or tri-parametric category of models, see Figure 2.

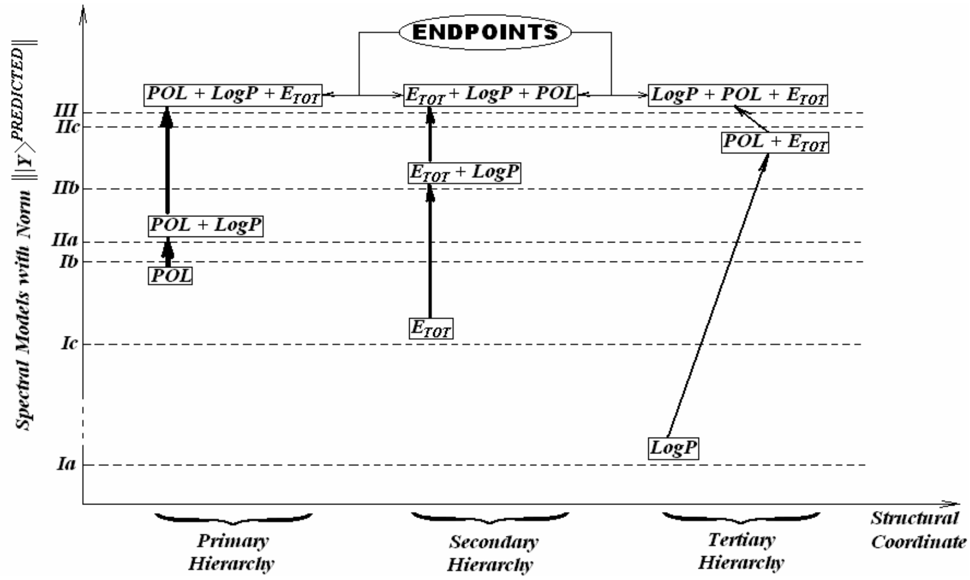

Figure 2. Spectral-structural models, designed through the rules of minimal spectral-SAR paths, emphasizing the primary, secondary and tertiary hierarchies forward the endpoints of the Tetrahymena pyriformis eco-toxicological activity [1].

Moreover, the logistic-spectral combined model was applied on ecotoxicological battery of organisms from selected species of Chlorella vulgaris, Vibrio fischeri and Pimephales promelas, giving the spectral representation of the molecular mechanisms intra- and interspecies based on the spectral paths and on their least principle. The resulted hierarchies (containing primary, secondary and tertiary molecular paths) are represented in the so called spectral scheme of interaction by the degree of thickness of the connecting lines, while the interference between the species toxicities are indicated by specific rectangular sign, for the primary hierarchies, and by circular sign for the secondary and tertiary hierarchies, see Figure 3.

Next, the ecotoxic-spectral-SAR model was accomplished for a series of ionic liquids activity measured on Vibrio fischeri species based on vectorial methodology generalization of structural data as the vectorial summation of cationic and anionic contributions in the resultant ionic liquid structure [6]

$$\left|Y_{AC}\right\rangle = \left|Y_{C}\right\rangle + \left|Y_{A}\right\rangle$$

while for the Daphnia magna action of the ionic liquids another scheme or anionic-cationic was employed in a generalized way [7]

$$\left|Y_{AC}\right\rangle^{0+} = \hat{O}_{S-SAR}\left|0+\right\rangle = \hat{O}_{S-SAR}\, f\left(\left\{\left|X_A\right\rangle\right\},\left\{\left|X_C\right\rangle\right\}\right)$$

with the specification of Hansch vectors:

$$f\left(LogP_A, LogP_C\right) \equiv LogP_{AC} = \log\left(e^{LogP_A} + e^{LogP_C}\right) \in \left\{\left|X_{1AC}\right\rangle\right\},$$

$$f\left(POL_A, POL_C\right) \equiv POL_{AC} = \left(POL_A^{1/3} + POL_C^{1/3}\right)^3 \in \left\{\left|X_{2AC}\right\rangle\right\} \text{ [Å}^3\text{]},$$

$$f\left(E_A, E_C\right) \equiv E_{AC} = E_A + E_C - 627.71\frac{q_A q_C}{POL_{AC}^{1/3}} \in \left\{\left|X_{3AC}\right\rangle\right\} \text{ [kcal/mol].}$$

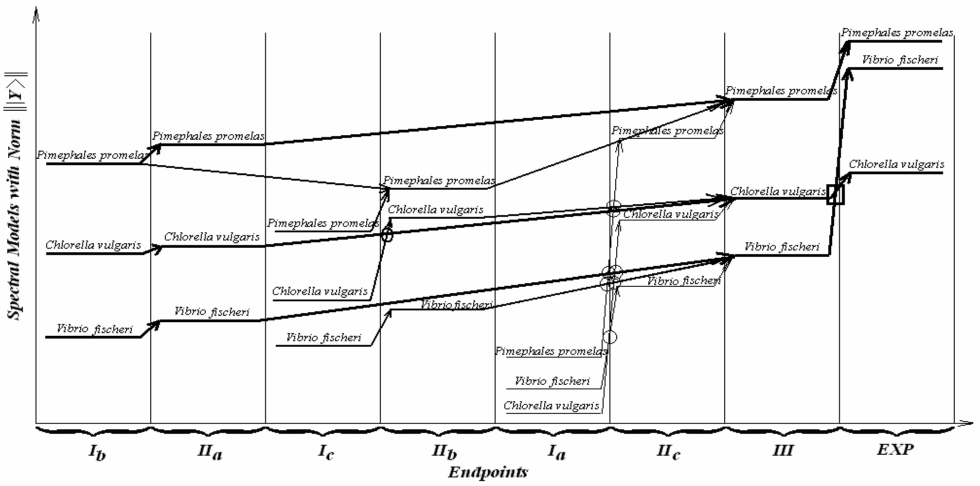

Figure 3. The spectral representation of the chemical-biological interaction paths across the S-SAR to experimental endpoints for the Chlorella vulgaris, Vibrio fischeri, and Pimephales promelas species, according with least path rule [5].

The obtained spectral hierarchies of the predicted activity relating different considered endpoints had provided the mechanistic scheme from which the cationic, anionic parts and ionic liquids in his integrity have different influence rules on the toxicity of the environment. There was concluded that while the tendencies of the cationic and anionic activities are complementary in a way they are not reduced in the ionic liquid summed up for approaching the total observed concentration in the spectral norm and correlation factor space, see Figure 4.

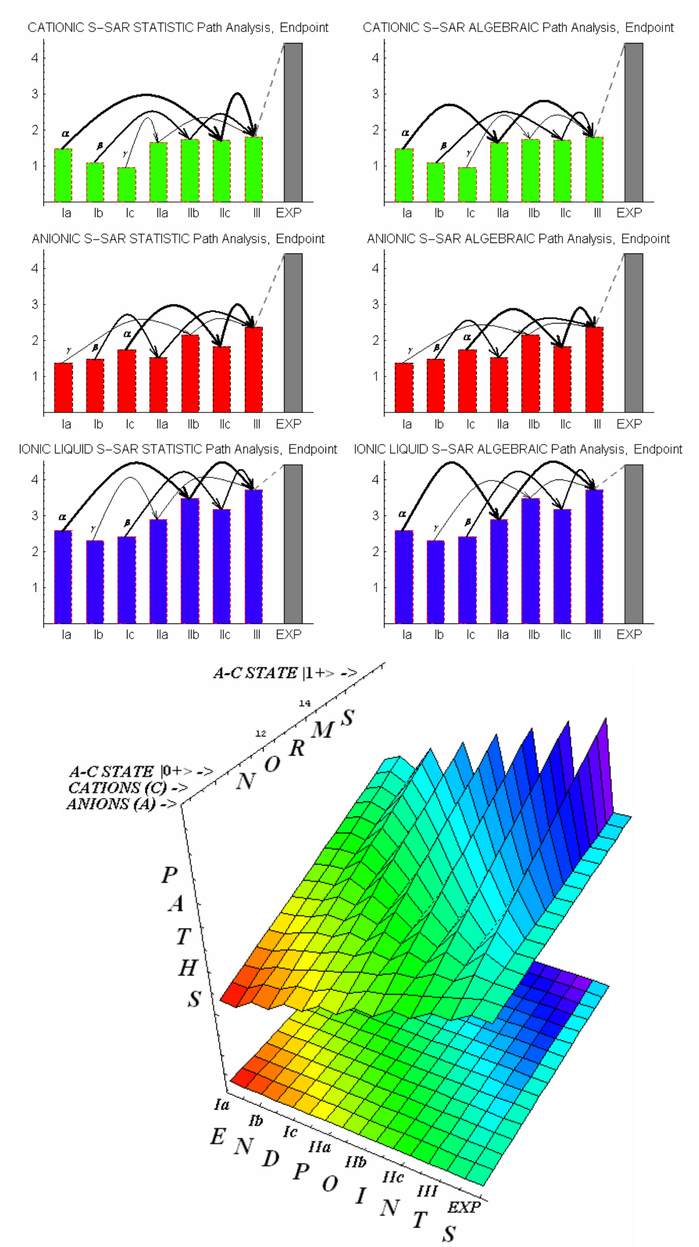

Figure 4. Spectral representation of the molecular hierarchical paths of the ecotoxicological action of specific ionic liquid series on Vibrio fischeri and Daphnia magna in left and right panels, respectively [6,7].

CONCLUSION

Through the structural and conceptual meaningful algorithm and the reliability in providing molecular paths of action in studied cases so far the present Spectral-SAR methodology opens the way in unifying the chemical-biological interactions and bonds in a 3D-dynamical (time dependency included) space.

REFERENCES

1. M.V. Putz, A.-M. Lacrămă, *Int. J. Molec. Sci*., **2007**, *8*, 363.
2. M.V. Putz, *Annals of West University of Timişoara, Series of Chemistry*, **2006**, *15*, 159.
3. M.V. Putz, A.-M. Lacrămă, *Annals of West University of Timişoara, Series of Chemistry,* **2006**, *15*, 167.
4. M.V. Putz, C. Duda-Seiman, D. M. Duda-Seiman, A.M. Putz, *Int. J. Chem. Modeling*, **2008**, *1*, on line (https://www.novapublishers.com).
5. A.-M. Lacrămă, M.V. Putz, V. Ostafe, *Advances in Quantum Chemical Bonding Structures*, Mihai V. Putz (Ed.), Transworld Research Network, Kerala, India **2008**, ISBN: 978-81-7895-306-9, Chapter 16, pp. 389-419.
6. A.-M. Lacrămă, M.V. Putz, V. Ostafe, *Int. J. Molec. Sci.*, **2007**, *8*, 842.
7. A.-M. Lacrămă, M.V. Putz, V. Ostafe, *Research Letter in Ecology*, ID12813/5 pages, DOI: 10.1155/2007/12813, **2007**.