

Article

Introducing Catastrophe-QSAR. Application on Modeling Molecular Mechanisms of Pyridinone Derivative-Type HIV Non-Nucleoside Reverse Transcriptase Inhibitors

Mihai V. Putz ^{1,*}, Marius Lazea ¹, Ana-Maria Putz ^{1,2,*} and Corina Duda-Seiman ¹

¹ Laboratory of Computational and Structural Physical Chemistry, Biology & Chemistry Department, West University of Timisoara, Str. Pestalozzi No. 16, 300115 Timisoara, Romania;

E-Mails: laz_marius2@yahoo.com (M.L.), cori_mam@yahoo.com (C.D.-S.)

² Institute of Chemistry Timisoara of the Romanian Academy, 24 Mihai Viteazul Bld, RO-300223, Timisoara, Romania

* Authors to whom correspondence should be addressed; E-Mails: mvputz@cbg.uvt.ro or mv_putz@yahoo.com (M.V.P.); putzanamaria@yahoo.com (A.M.P.); Tel.: +40-256-592-633; Fax: +40-256-592-620.

Received: 27 October 2011; in revised form: 28 November 2011 / Accepted: 12 December 2011 /

Published: 20 December 2011

Abstract: The classical method of quantitative structure-activity relationships (QSAR) is enriched using non-linear models, as Thom's polynomials allow either uni- or bi-variate structural parameters. In this context, catastrophe QSAR algorithms are applied to the anti-HIV-1 activity of pyridinone derivatives. This requires calculation of the so-called relative statistical power and of its minimum principle in various QSAR models. A new index, known as a statistical relative power, is constructed as an Euclidian measure for the combined ratio of the Pearson correlation to algebraic correlation, with normalized t-Student and the Fisher tests. First and second order inter-model paths are considered for mono-variate catastrophes, whereas for bi-variate catastrophes the direct minimum path is provided, allowing the QSAR models to be tested for predictive purposes. At this stage, the max-to-min hierarchies of the tested models allow the interaction mechanism to be identified using structural parameter succession and the typical catastrophes involved. Minimized differences between these catastrophe models in the common structurally influential domains that span both the trial and tested compounds identify the "optimal molecular structural domains" and the molecules with the best output with respect to the modeled activity, which in this case is human immunodeficiency virus type 1 HIV-1

inhibition. The best molecules are characterized by hydrophobic interactions with the HIV-1 p66 subunit protein, and they concur with those identified in other 3D-QSAR analyses. Moreover, the importance of aromatic ring stacking interactions for increasing the binding affinity of the inhibitor-reverse transcriptase ligand-substrate complex is highlighted.

Keywords: Thom's catastrophe polynomials; statistical factors; minimum statistical paths; QSAR structural domains; HIV-1 inhibitory activity

1. Introduction

Among the mathematical theories that model open-system dynamics, Thom's theory of catastrophes has acquired much popularity for its simple yet valuable description of the system-environment interaction that includes phenomena such as steady state equilibrium and life cycles [1]. In particular, biological systems come first under catastrophe modeling because they display a causal action-reaction response to various natural or imposed constraining limits. As an example, the reactions of organisms to vital toxicological threats were developed into the survival attractor concept by employing butterfly bifurcation phenomenology, which is closely related to the cusp catastrophe, thus revealing the close connection with the turning points around singularity points of the fundamental central field laws of attraction [2]. The cusp catastrophe was further implemented in the physiological processes of predation and generation, thus giving mathematical support to Heidegger's philosophical concept of *entity* and having the major consequence of translating the ontological entities into computer language [3]. Following this line of application, Jungian psychology entered the topological approach phase through modeling personal unconscious and conscious states using the swallowtail catastrophe [4]. As a consequence, neuro-self-organization was advanced by reduction to cusp synergetics as an archetypal precursor of epileptic seizures [5]. Nevertheless, in chemistry the catastrophe approach enters through the need to unitarily characterize elementary processes such as chemical bonding, leading to the so-called bonding evolution theory and reformulation of the electronic localization functions [6,7]. In the last decade, catastrophe theory was successfully grounded on Hilbert space modeling with the density matrix and non-linear evolution as specific tools for the non-commutative (quantum) systems [8]. At this point, the interesting connection with the linear superposition of quantum states may be generalized in a non-linear manner with direct correspondence for widespread quantitative structure-activity relationship (QSAR) treatments of the "birth and death of an organism".

In this context, the present contribution provides *in silico* assistance to clinical efforts in current antiretroviral therapy by contributing to the development of a given class of actual anti-HIV-1 compounds and identifying their viral inhibitory mechanisms and influential structural factors. Continuous efforts both in theory and in clinical practice are made to provide new and valid data for HIV infection management. Note that acquired immunodeficiency deficiency syndrome (AIDS) was first recognized in 1981. Only 25 compounds have been approved for use in HIV infected patients, and they are distributed among several classes of antiretroviral drug types [9,10]: nucleoside reverse transcriptase inhibitors (NRTIs); nucleotide reverse transcriptase inhibitors (NtRTIs); non-nucleoside

reverse transcriptase inhibitors (NNRTIs); protease inhibitors (PIs); cell entry (or fusion) inhibitors (FIs); co-receptor inhibitors (CRIs); and integrase inhibitors (INIs). Among these, it is well known that most NNRTIs have a low genetic barrier to resistance, *i.e.*, high viral resistance may be induced by a single mutation at the NNRTI binding site [11]. It is this particular feature that makes NNRTIs so well adapted for a comprehensive catastrophe theory application. Although NNRTIs are an open battlefield for research, being highly active in naïve and drug-resistant HIV infected patients [12], QSAR methods are cost-effective approaches to developing new and potent molecules with increased anti-HIV activity [13–23]. As a viable alternative to the available 3D-QSARs, the present endeavor makes the first steps toward generalizing multi-linear QSAR to non-linear catastrophe QSAR analysis and toward providing a conceptual-computational framework in which both the interactions occurring between the pyridinone derivatives and the NNRTI binding site and the structural domains influential for HIV-1 RT inhibitory activity are accounted for [24].

2. Background Theories

2.1. QSAR Phenomenology

The fundamental problem of structure-activity analysis may be described as follows: given a congener set of N -compounds/molecules with measured/observed activity (A) one searches for the best correlation of it with the structural (intrinsic, internal) molecular information quantified by M -properties (such as hydrophobicity, polarization, total energy), classically presented in multi-linear form [25–31]:

$$Y = b_0 + b_1 X_1 + \dots + b_k X_k + \dots + b_M X_M \quad (1)$$

Equation (1) has some basic features, namely:

- Y stands for the computed activity, not the observed activity, from the statistical characteristics of the present approach; thus the validation of Equation (1) should be done for another (preferably external or testing) set of compounds with which the predictive power of Equation (1) is tested.
- Because the right side of Equation (1) unfolds as a linear summation of the structural characteristics, it corresponds in fact with the quantum superposition principle, which provides a global Eigen-solution for a quantum system from its particular realization in orthogonal or projective sub-space; from where the need arises for structural indices X_1, \dots, X_M to be either linearly independent or orthogonal in algebraic space built from their associate vectors presented in Table 1.

Table 1. The QSAR working table for Equation (1) in the presence of M -structural descriptors for N -compounds with known activities.

Observed Activity	Structural Predictor Variables				
A	X_1	...	X_k	...	X_M
A_1	x_{11}	...	x_{1k}	...	x_{1M}
A_2	x_{21}	...	x_{2k}	...	x_{2M}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
A_N	x_{N1}	...	x_{Nk}	...	x_{NM}

However, in order for the chemical structure be correlated with bio-, eco-, or pharmacological activity in an analytical manner (from where the name Quantitative Structure-Activity Relationship arises) that has sense for the ligand-receptor interaction under study, the Organization for Economic Cooperation and Development (OECD) developed the so-called QSAR-OECD principles, which have already been adopted by the EU Parliament as the official guidelines for further regulation of compounds in the European Union. They are, in short [32]:

- QSAR 1: a defined endpoint
- QSAR-2: an unambiguous algorithm
- QSAR-3: a defined domain of applicability
- QSAR-4: appropriate measures of goodness-of-fit, robustness and predictivity
- QSAR-5: a mechanistic interpretation, if possible

Put differently, they express the essence of the chemical modeling of biological effects while relaying (Husserl-Russell) knowledge phenomenology in a more general manner [33]:

- QSAR-1. why does one do modeling ?
- QSAR-2. how does one do modeling ?
- QSAR-3. with what tools do I model ?
- QSAR-4. how reliable is what I modeled ?
- QSAR-5. what knowledge did the model provide ?

Therefore, although the backbone of QSAR modeling is based on equation (1), one should be aware that it represents, despite the innumerable extant studies, only one type of model—the multi-linear type. It is therefore worth refreshing QSAR studies by exploring other ways of combining the structural parameters that cause the observed biological activity. However, although it is clear that non-linear QSAR is the next generation of correlations, one should not search arbitrarily or randomly while having at hand a well-designed theory of non-linear modeling of natural phenomena: Thom's catastrophe theory, the basic assumptions and main working tools of which are presented next.

2.2. Thom's Catastrophe Theory

René Thom's catastrophe theory basically describes how, for a given system, continuous action on the *control space* (C^k), parameterized by C_k 's, provides a sudden change in its *behavior space* (I^m), described by x_m variables through stable singularities of the smooth map [34,35]

$$\eta(c_k, x_m): C^k \times I^m \rightarrow \mathfrak{R} \quad (2)$$

with $\eta(c_k, x_m)$ called the *generic potential* of the system. Therefore, catastrophes are given by the set of *critical points* (c_k, x_m) for which the field gradient of the generic potential vanishes

$$M^{k \times m} = \left\{ (c_k, x_m) \in C^k \times I^m \mid \nabla_{x_m} \eta(c_k, x_m) = 0 \right\} \quad (3)$$

or, more rigorously: a catastrophe is a singularity of the map $M^{k \times m} \rightarrow C^k$.

Next, depending on the number of parameters in space C^k (also called the *co-dimension*, k) and on the number of variables in space I^m (also called the *co-rank*, m), Thom classified the generic potentials (or maps) given by Equation (2) as seven unfolding elementary (in the sense of universal)

catastrophes, *i.e.*, providing the multi-variable (with the co-rank up to two) and multi-parametrical (with the co-dimension up to four) polynomials listed in Table 2. Going to the higher derivatives of the generic potential (the fields), the control parameter c_k^* for which the Laplacian of the generic potential vanishes

$$\Delta_x \eta(c_k^*, x_m) = 0 \tag{4}$$

gives the *bifurcation point*. Consequently, the set of control parameters $c^\#$ for which the Laplacian of a critical point is non-zero defines the *domain of stability* of the critical point. It is clear now that small perturbations of $\eta(c^*, x)$ bring the system from one domain of stability to another; otherwise, the system is located within a *domain of structural stability*.

Table 2. Thom’s Classification of Elementary Catastrophes [36,37].

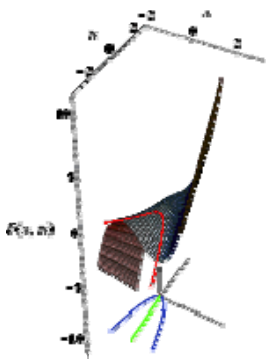
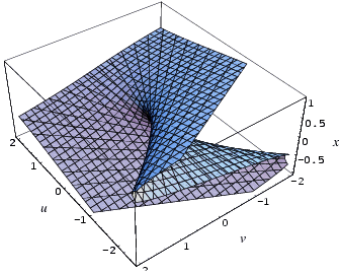
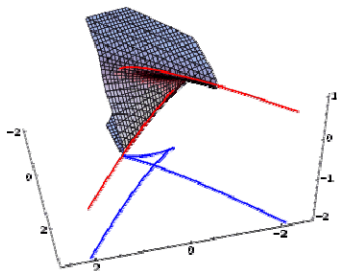
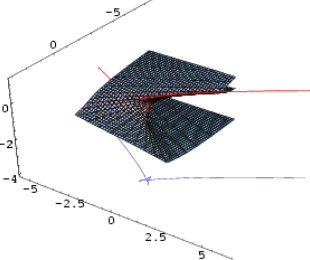
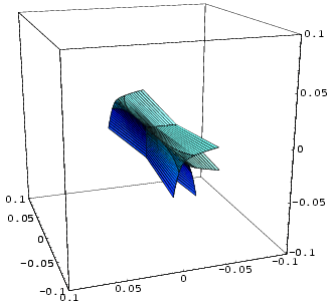
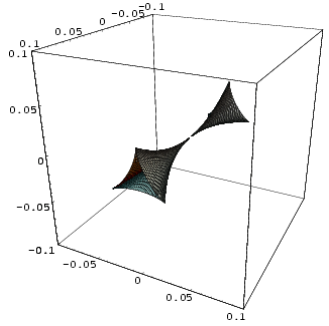
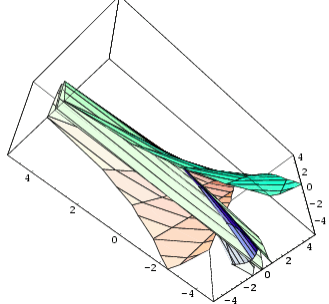
Name	Co-dimension	Co-rank	Universal unfolding	Parametric Representation
Fold	1	1	$x^3 + ux$	
Cusp	2	1	$x^4 + ux^2 + vx$	
Swallow tail	3	1	$x^5 + ux^3 + vx^2 + wx$	
Butterfly	4	1	$x^6 + ux^4 + vx^3 + wx^2 + tx$	

Table 2. Cont.

Hyperbolic umbilic	3	2	$x^3 + y^3 + uxy + vx + wy$	
Elliptic umbilic	3	2	$x^3 - xy^2 + u(x^2 + y^2) + vx + wy$	
Parabolic umbilic	4	2	$x^2y + y^4 + ux^2 + vy^2 + wx + ty$	

Remarkably, the cases described above correspond to the equilibrium limit of the dynamical (non-equilibrium) evolution of an open system

$$F\left(c_k; t; \eta(c_k; x_m); \frac{\partial \eta(c_k; x_m)}{\partial t}, \dots\right) = 0 \tag{5}$$

where the behavior space is further parameterized by the temporal paths $x_m(c_k, t)$. The connection with equilibrium is recovered through the stationary time regime imposed on the critical points. In this way, the set of points giving a critical point in the stationary $t \rightarrow +\infty$ regime (the so-called ω -limit) corresponds to an attractor, and it forms a basin, whereas the stationary regime $t \rightarrow -\infty$ (the so-called α -limit) describes a repeller. In this way, the catastrophe polynomials may be regarded either as an asymptotic solution of a dynamical evolutionary system or as a steady state solution allowing the quasi-equilibrium of the ligand-receptor or inhibitor-organism interactions to be described. However, in complex binding systems with multiple evolutionary phases, e.g., the HIV-1 life cycle, the possibility of “linking” the various classes of catastrophes themselves may provide a striking analytical approach to the dynamics and mutational sensitivity of the studied interaction that starts with the actual catastrophe-QSAR method.

3. Catastrophe-QSAR Method

Aiming to construct a QSAR rationale from the elementary catastrophes, the next steps are implemented:

- i. Assuming the vectorial form of activities and of associated QSARs are according to Table 2, Table 3 showing catastrophe-QSAR is thereby formed.

Table 3. Algebraic realization of Thom’s elementary catastrophes as uni- and bi- nonlinear QSARs. The systematics of the sub-indices indicate consecutive coupled pairs, where each pair is interpreted as: the index of a structural factor followed by its power.

Model	QSAR Equation
GROUP I: with one descriptor only, $ X_1\rangle$	
QSAR-(I)	$ Y_I\rangle = a_0 1\rangle + a_{11} X_1\rangle$
Fold	$ Y_F\rangle = f_0 1\rangle + f_{11} X_1\rangle + f_{13} X_1^3\rangle$
Cusp	$ Y_C\rangle = c_0 1\rangle + c_{11} X_1\rangle + c_{12} X_1^2\rangle + c_{14} X_1^4\rangle$
Swallow tail	$ Y_{ST}\rangle = s_0 1\rangle + s_{11} X_1\rangle + s_{12} X_1^2\rangle + s_{13} X_1^3\rangle + s_{15} X_1^5\rangle$
Butterfly	$ Y_B\rangle = b_0 1\rangle + b_{11} X_1\rangle + b_{12} X_1^2\rangle + b_{13} X_1^3\rangle + b_{14} X_1^4\rangle + b_{16} X_1^6\rangle$
GROUP II: with two descriptors, $ X_1\rangle, X_2\rangle$	
QSAR- (II)	$ Y_{II}\rangle = q_0 1\rangle + q_{11} X_1\rangle + q_{21} X_2\rangle$
Hyperbolic umbilic	$ Y_{HU}\rangle = h_0 1\rangle + h_{11} X_1\rangle + h_{21} X_2\rangle + h_{1121} X_1X_2\rangle + h_{13} X_1^3\rangle + h_{23} X_2^3\rangle$
Elliptic umbilic	$ Y_{EU}\rangle = e_0 1\rangle + e_{11} X_1\rangle + e_{21} X_2\rangle + e_{12} X_1^2\rangle + e_{22} X_2^2\rangle + e_{1122} X_1X_2^2\rangle + e_{13} X_1^3\rangle$
Parabolic umbilic	$ Y_{PU}\rangle = p_0 1\rangle + p_{11} X_1\rangle + p_{21} X_2\rangle + p_{12} X_1^2\rangle + p_{22} X_2^2\rangle + p_{1221} X_1^2X_2\rangle + p_{24} X_2^4\rangle$

- ii. Determine the norms for each model

$$\|Y\rangle\| = \sqrt{\langle Y|Y\rangle} = \sqrt{\sum_{i=1}^N y_i^2} \tag{6}$$

- iii. Calculate the algebraic correlation factor for each model [31]

$$R_{ALG} = \frac{\|Y\rangle\|}{\|A\rangle\|} = \sqrt{\frac{\sum_{i=1}^N y_i^2}{\sum_{i=1}^N A_i^2}} \tag{7}$$

- iv. Calculate the so-called “statistical relative power” index for each model with each set of descriptors

$$\Pi = \sqrt{r^2 + t^2 + f^2} \tag{8}$$

where the components are defined as follows:

- relative index of correlation:

$$r = \frac{R_{ALG}}{R_{Pearson}} \tag{9}$$

- relative index for Student's *t*-test

$$t = \frac{t_{Computed}}{t_{Tabulated}^{(1-\alpha=0.99; N-M-2)}} \tag{10}$$

- relative index for Fisher's test

$$f = \frac{F_{Computed}}{F_{Tabulated}^{(1-\alpha=0.99; M, N-M-1)}} \tag{11}$$

- v. Determine the generalized Euclidian distances between corresponding type-I and type-II models employing different descriptors

$$\Delta\Pi = \sqrt{(r - r')^2 + (t - t')^2 + (f - f')^2} \tag{12}$$

and establish formal matrices for the models' differences for single descriptors, respectively

$$\Delta^2\Pi_{I(X_1, X_2)} = \left| \Delta\Pi_{I(X_1)} - \Delta\Pi_{I(X_2)} \right| \tag{13}$$

where

$$\Delta\Pi_{I(X=X_1 \vee X_2)} = \begin{pmatrix} QSAR_{I(X)} - F_{(X)} & QSAR_{I(X)} - C_{(X)} & QSAR_{I(X)} - ST_{(X)} & QSAR_{I(X)} - B_{(X)} \\ & F_{(X)} - C_{(X)} & F_{(X)} - ST_{(X)} & F_{(X)} - B_{(X)} \\ & & C_{(X)} - ST_{(X)} & C_{(X)} - B_{(X)} \\ & & & ST_{(X)} - B_{(X)} \end{pmatrix} \tag{14}$$

and for pair descriptors

$$\Delta\Pi_{II(X_1 \wedge X_2)} = \begin{pmatrix} QSAR_{II(X_1, X_2)} - HU_{(X_1, X_2)} & QSAR_{II(X_1, X_2)} - EU_{(X_1, X_2)} & QSAR_{II(X_1, X_2)} - PU_{(X_1, X_2)} \\ & HU_{(X_1, X_2)} - EU_{(X_1, X_2)} & HU_{(X_1, X_2)} - PU_{(X_1, X_2)} \\ & & EU_{(X_1, X_2)} - PU_{(X_1, X_2)} \end{pmatrix} \tag{15}$$

- vi. Identify all minimum paths across all differences $\Delta\Pi_{I(X_1 \vee X_2)}$, $\Delta^2\Pi_{I(X_1, X_2)}$ and $\Delta\Pi_{II(X_1 \wedge X_2)}$ for a given set of descriptors (X_1, X_2)

$$\begin{cases} \delta\{\Delta\Pi_{I(X)}\} = 0 \\ \delta\{\Delta^2\Pi_{I(X_1 \vee X_2)}\} = 0 \\ \delta\{\Delta\Pi_{II(X_1 \wedge X_2)}\} = 0 \end{cases} \tag{16}$$

The combination of descriptors that fulfills this system provides the *molecular mechanism* of the interaction. The correlation models involved are ordered according to their relative statistical power within the same molecular mechanism, thereby providing the *best models*. Because pair-descriptors are primarily involved in the present analysis, one can consider the first two such “waves” and their best correlation models up to the second order minimum paths, as in Equation (16).

vii. For selected correlation models, in either structure-driven or molecular mechanistic “waves,” one employs them to compute the associated predicted activities for test molecules and to provide the statistics regarding the observed activity. If the obtained relative statistical power is close to those characteristic for the trial set of molecules, then these models may be validated for the specific eco-, bio-, or pharmacological problem. Moreover, further insight will be provided by the analysis of the catastrophe shape of the models involved and discussed accordingly.

Nevertheless, more Catastrophe Theory insights and the natural consequence on statistical (Pearson) correlation behavior may be found in Appendix.

4. Application to Non-Nucleoside Reverse Transcriptase Pyridinone Inhibitors

4.1. Input Data

As a working molecular series, the interesting series of pyridinone derivatives in Table 4 is herein employed [24] because of their potential for improving and complementing the currently available four NNRTIs that have been approved by the U.S. FDA for HIV/AIDS treatment (Nevirapine-Viramune[®], Delavirdine-Rescriptor[®], Efavirenz-Sustiva[®], Etravirine-Intelence[®]), all of which bind to the hydrophobic pocket of HIV-1 reverse transcriptase [38]. The pyridinone derivatives were divided into a training set of 23 compounds and a test set of 9 compounds according to the methods of normal/Gaussian (G) and non-normal/non-Gaussian (NG) fitted activity [39–41] (Figure 1).

Figure 1. Gaussian (G) and non-Gaussian (NG) screening of the observed activities of the working molecules in Table 4 grouped into trial and test congener series.

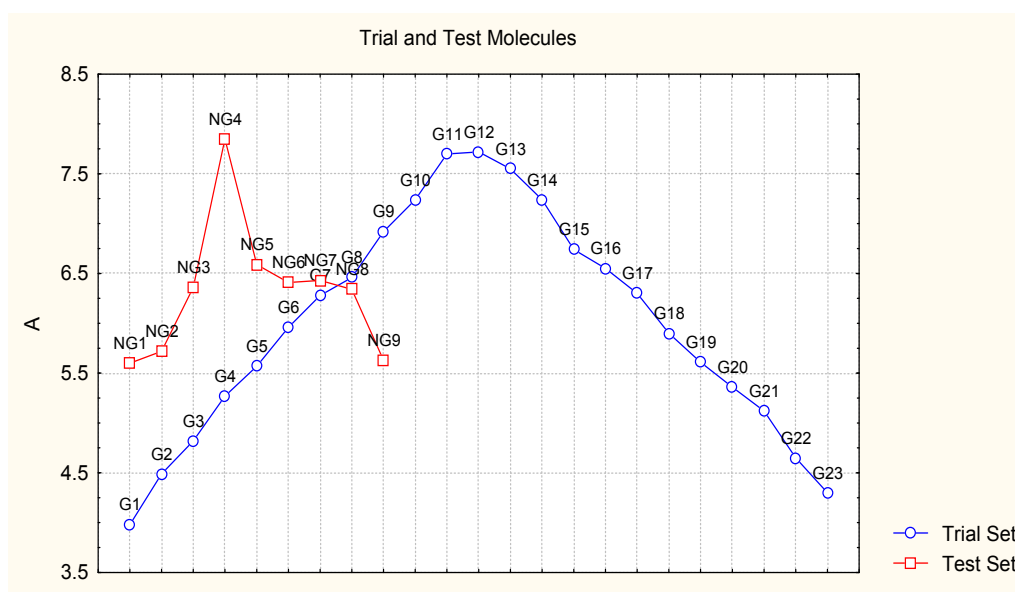


Table 4. Actual working reverse transcriptase pyridinone inhibitors grouped in Gaussian (G) and non-Gaussian (NG) molecular congeneric sets with their structural information (hydrophobicity, Log P; molecular polarizability POL [\AA^3] and total optimized energy of formation H [kcal/mol]) computed upon the semi-empirical PM3 method [42], along with their observed activity $A = \text{Log}(1/\text{IC}_{50})$ [24].

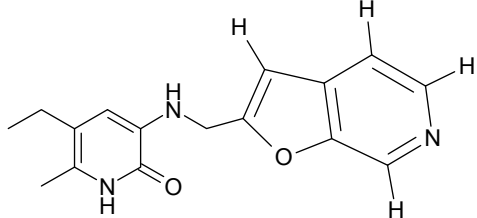
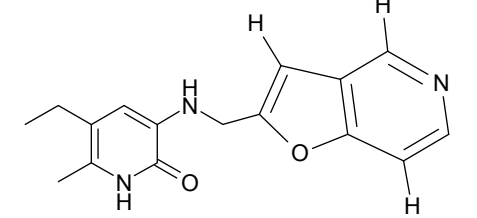
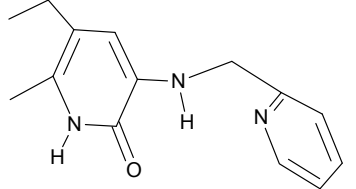
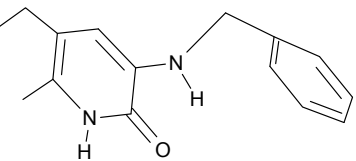
No.	Type	WORKING MOLECULES		A^{obs} Log (1/IC ₅₀)	QSAR parameters		
		Structure	Name		Log P	POL (\AA^3)	H (kcal/mol)
1.	G1		3-[[[(6'-azabenzofuran-2'-yl) methyl]amino]-5-ethyl-6-methylpyridin-2(1H)-one	3.98	-0.54	31.21	-14.67
2.	G2		3-[[[(5'-azabenzofuran-2'-yl) methyl]amino]-5-ethyl-6-methylpyridin-2(1H)-one	4.49	-0.54	31.21	-16.195
3.	G3		3-[[[(pyridine-2'-yl) methyl]amino]-5-ethyl-6-methylpyridin-2(1H)-one	4.82	0.21	27.87	-5.854
4.	G4		3-benzylamino-5-ethyl-6-methylpyridin-2(1H)-one	5.27	0.67	28.58	-11.659

Table 4. Cont.

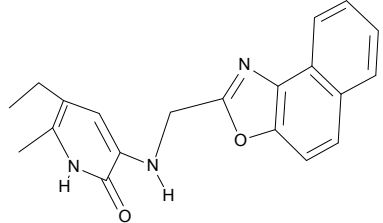
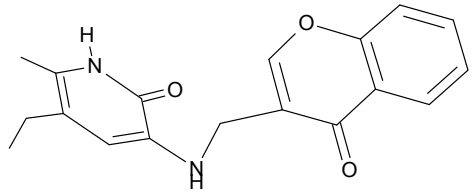
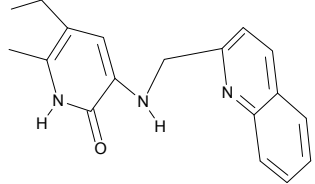
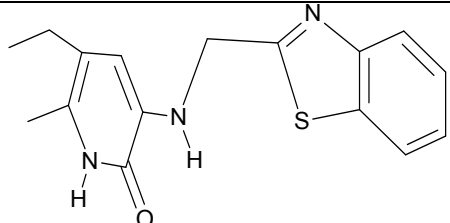
No.	Type	WORKING MOLECULES		A ^{obs} Log (1/IC ₅₀)	QSAR parameters		
		Structure	Name		Log P	POL (Å ³)	H (kcal/mol)
5.	G5		3-[[1',3'-naftoxazol-2'-yl)methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	5.57	1.20	38.48	-1.878
6.	G6		3-[[1'-benzopyran-4'-one-3'-yl)methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	5.96	-0.71	33.84	-61.455
7.	G7		3-[[benzopyridine-2'-yl)methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.28	1.16	35.14	11.246
8.	G8		3-[[1',3'-benzothiazole-2'-yl)methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.46	0.54	33.57	17.808

Table 4. Cont.

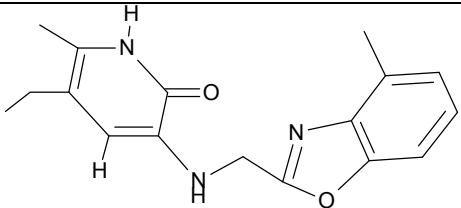
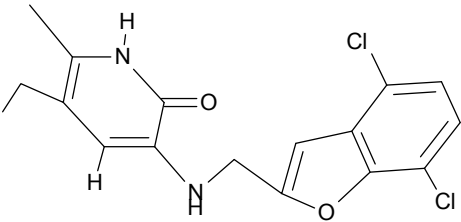
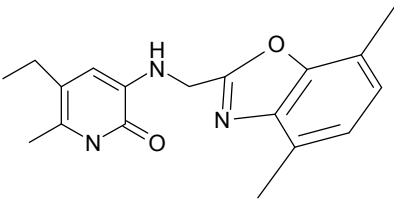
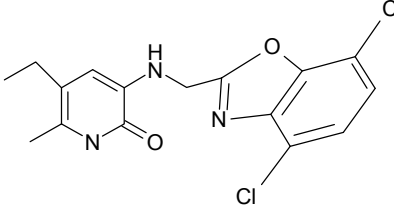
No.	Type	WORKING MOLECULES		A ^{obs}	QSAR parameters		
		Structure	Name		Log (1/IC ₅₀)	Log P	POL (Å ³)
9.	G9		3-[[4'-methyl-benzoxazole-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.92	0.67	33.05	-27.613
10.	G10		3-[[4',7'-dichloro-benzofuran-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	7.24	0.88	35.78	-33.749
11.	G11		3-[[4',7'-dimethyl-benzoxazol-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	7.7	1.13	34.88	-38.048
12.	G12		3-[[4',7'-dichloro-benzoxazol-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	7.72	1.24	35.07	-30.071

Table 4. Cont.

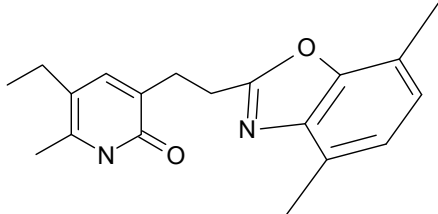
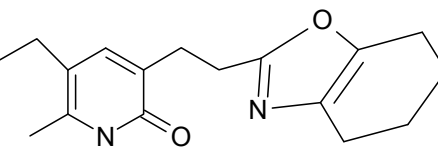
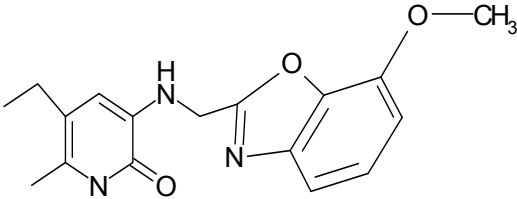
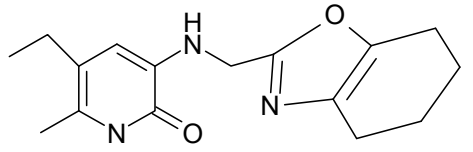
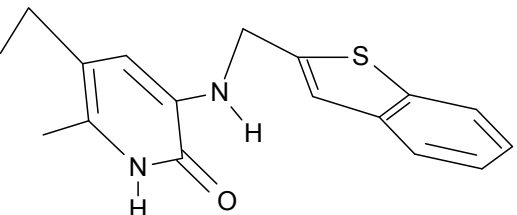
No.	Type	WORKING MOLECULES		A ^{obs}	QSAR parameters		
		Structure	Name	Log (1/IC ₅₀)	Log P	POL (Å ³)	H (kcal/mol)
13.	G13		3-[(4',7'-dimethylbenzoxazol-2'-yl) ethyl]-5-ethyl-6-methylpyridin-2(1H)-one	7.55	2.62	35.37	-47.701
14.	G14		3-[(4',5',6',7'-tetrahydrobenzoxazole-2'-yl) ethyl]-5-ethyl-6-methylpyridin-2(1H)-one	7.24	-0.02	32.08	-63.299
15.	G15		3-[(4'-methoxybenzoxazole-2'-yl) methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.74	-0.05	33.68	-54.452
16.	G16		3-[(4',5',6',7'-tetrahydrobenzoxazole-2'-yl) methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.55	-1.50	31.59	-50.643
17.	G17		3-[(benzothiophene-2'-yl) methyl] amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.30	0.19	34.28	11.703

Table 4. Cont.

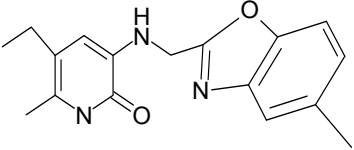
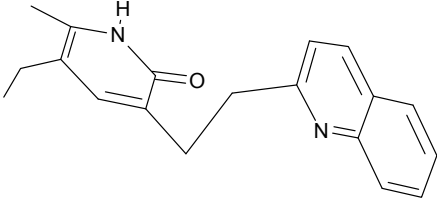
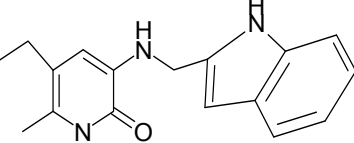
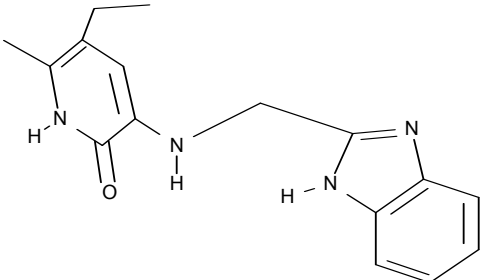
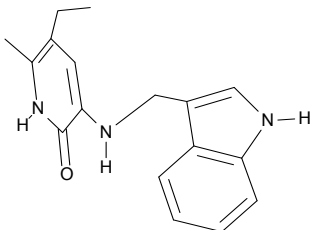
No.	Type	WORKING MOLECULES		A ^{obs}	QSAR parameters		
		Structure	Name	Log (1/IC ₅₀)	Log P	POL (Å ³)	H (kcal/mol)
18.	G18		3-[[5-(5-ethyl-6-methylpyridin-2(1H)-one)methyl]amino]benzoxazole	5.90	0.67	33.05	-27.741
19.	G19		3-[[5-(5-ethyl-6-methylpyridin-2(1H)-one)ethyl]benzopyridine	5.61	2.71	35.62	3.331
20.	G20		3-[[5-(5-ethyl-6-methylpyridin-2(1H)-one)methyl]amino]indole	5.36	-0.34	32.63	4.727
21.	G21		3-[[5-(5-ethyl-6-methylpyridin-2(1H)-one)methyl]amino]quinazolin-2(1H)-one	5.12	0.02	31.92	8.171
22.	G22		3-[[5-(5-ethyl-6-methylpyridin-2(1H)-one)methyl]amino]indole	4.65	-0.43	32.63	2.957

Table 4. Cont.

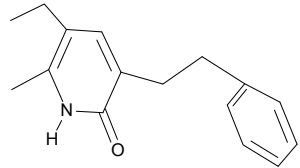
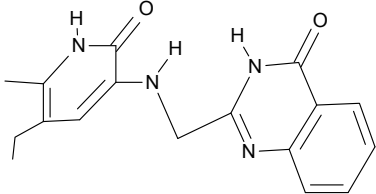
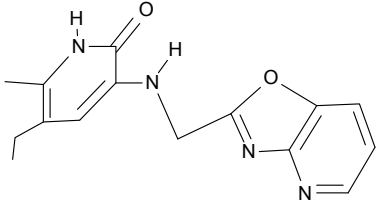
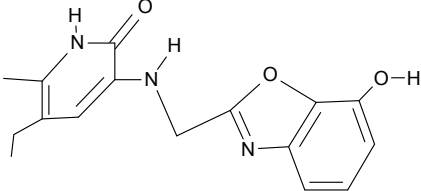
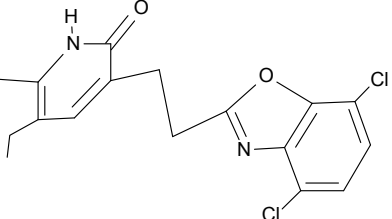
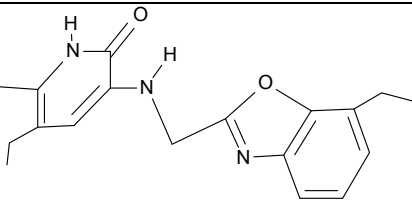
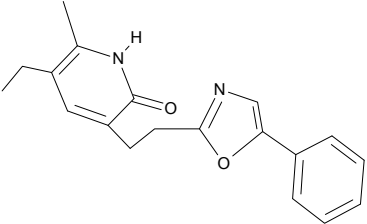
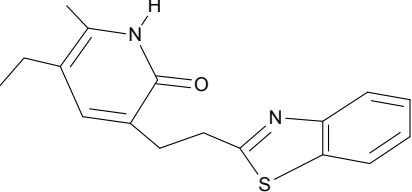
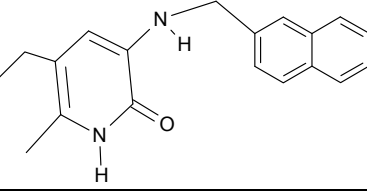
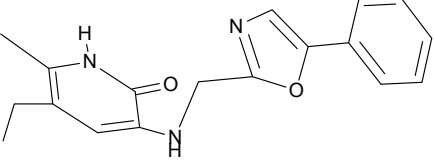
No.	Type	WORKING MOLECULES		A ^{obs}	QSAR parameters		
		Structure	Name	Log (1/IC ₅₀)	Log P	POL (Å ³)	H (kcal/mol)
23.	G23		3-(β-phenylethyl)-5-ethyl-6-methylpyridin-2(1H)-one	4.30	2.36	29.06	-23.245
24.	NG1		3-[[4'-quinoxalone-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	5.60	-0.47	33.85	-36.959
25.	NG2		3-[[3',4'-diazobenzofuran-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	5.72	0.05	30.50	-8.120
26.	NG3		3-[[7'-hydroxybenzoxazole-2'-yl]methyl]amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.36	-0.08	31.85	-62.189
27.	NG4		3-[(4',7'-dichlorobenzoxazole-2'-yl) ethyl]-5-ethyl-6-methylpyridin-2(1H)-one	7.85	2.72	35.55	-39.459

Table 4. Cont.

No.	Type	WORKING MOLECULES		A ^{obs} Log (1/IC ₅₀)	QSAR parameters		
		Structure	Name		Log P	POL (Å ³)	H (kcal/mol)
28.	NG5		3-[[7'-ethyl-methylamino}-5-ethyl-6-methylpyridin-2(1H)-one	6.59	1.06	34.88	-34.478
29.	NG6		3-[(5'-phenyl-oxazole-2'-yl) ethyl]-5-ethyl-6-methylpyridin-2(1H)-one	6.41	0.96	35.17	-21.361
30.	NG7		3-[(benzothiazole-2'-yl) ethyl]-5-ethyl-6-methylpyridin-2(1H)-one	6.43	2.02	34.06	8.873
31.	NG8		3-[[2'-naphthyl] methyl] amino}-5-ethyl-6-methylpyridin-2(1H)-one	6.34	1.67	35.85	5.495
32.	NG9		3-[[5'-phenyl-oxazole-2'-yl] methyl] amino}-5-ethyl-6-methylpyridin-2(1H)-one	5.63	-0.53	34.69	-10.850

4.2. Results and Discussion

The catastrophe-QSAR algorithm of Section 3 was applied to the molecules of Table 4, and the trial results are presented in Tables 5–9.

Table 5. Correlation equations for the Group-I models of Table 3 and the molecular structures and data of Table 4.

Catastrophe	QSAR Model	$R_{\text{Pearson}}^{(a)}$	$R_{\text{ALG}}^{(b)}$	$r^{(c)}$	$t\text{-Stud.}$	$t^{(d)}$	Fisher	$f^{(e)}$	$\Pi^{(f)}$
QSAR (I)	$ Y_I^{\text{LogP}}\rangle = 5.861 1\rangle + 0.240 \text{LogP}\rangle$	0.228	0.984	4.317	22.344	7.854	1.150	0.143	8.963
	$ Y_I^{\text{POL}}\rangle = -2.257 1\rangle + 0.249 \text{POL}\rangle$	0.554	0.989	1.784	-0.832	-0.292	9.284	1.158	2.147
	$ Y_I^H\rangle = 5.57 1\rangle - 0.021 H\rangle$	0.476	0.987	2.074	20.597	7.24	6.156	0.768	7.57
Fold (F)	$ Y_F^{\text{LogP}}\rangle = 5.854 1\rangle + 0.738 \text{LogP}\rangle - 0.106 \text{LogP}^3\rangle$	0.382	0.986	2.581	22.936	8.062	1.705	0.213	8.468
	$ Y_F^{\text{POL}}\rangle = -24.206 1\rangle + 1.26 \text{POL}\rangle - 3 \cdot 10^{-4} \text{POL}^3\rangle$	0.601	0.989	1.646	-1.422	-0.45	5.650	0.704	1.859
	$ Y_F^H\rangle = 5.58 1\rangle - 0.016 H\rangle - 2 \cdot 10^{-6} H^3\rangle$	0.481	0.987	2.053	20.095	7.063	3.01	0.375	7.365
Cusp (C)	$ Y_C^{\text{LogP}}\rangle = 5.707 1\rangle + 0.426 \text{LogP}\rangle + 0.372 \text{LogP}^2\rangle - 0.071 \text{LogP}^4\rangle$	0.348	0.985	2.832	16.120	5.666	0.872	0.109	6.335
	$ Y_C^{\text{POL}}\rangle = 431.26 1\rangle - 35.694 \text{POL}\rangle + 0.833 \text{POL}^2\rangle - 10^{-4} \text{POL}^4\rangle$	0.713	0.992	1.391	2.240	0.787	6.558	0.818	1.796
	$ Y_C^H\rangle = 5.006 1\rangle + 0.042 H\rangle + 0.003 H^2\rangle - 10^{-6} H^4\rangle$	0.764	0.993	1.300	19.802	6.960	8.864	1.105	7.166
Swallow tail (ST)	$ Y_{ST}^{\text{LogP}}\rangle = 5.649 1\rangle + 1.608 \text{LogP}\rangle + 0.326 \text{LogP}^2\rangle - 0.978 \text{LogP}^3\rangle + 0.093 \text{LogP}^5\rangle$	0.575	0.989	1.720	18.665	6.561	2.222	0.277	6.788
	$ Y_{ST}^{\text{POL}}\rangle = 1476.244 1\rangle - 156.079 \text{POL}\rangle + 5.791 \text{POL}^2\rangle - 0.079 \text{POL}^3\rangle + 5.5 \cdot 10^{-6} \text{POL}^5\rangle$	0.715	0.992	1.387	0.45	0.158	4.708	0.587	1.515
	$ Y_{ST}^H\rangle = 4.884 1\rangle + 0.031 H\rangle + 0.004 H^2\rangle + 5.2 \cdot 10^{-5} H^3\rangle + 4 \cdot 10^{-10} H^5\rangle$	0.763	0.993	1.302	15.608	5.486	6.263	0.781	5.692

Table 5. Cont.

Catastrophe	QSAR Model	$R_{\text{Pearson}}^{(a)}$	$R_{\text{ALG}}^{(b)}$	$r^{(c)}$	$t\text{-Stud.}$	$t^{(d)}$	Fisher	$f^{(e)}$	$\Pi^{(f)}$
Butterfly (B)	$ Y_B^{\text{LogP}}\rangle = 5.646 1\rangle + 1.464 \text{LogP}\rangle + 0.303 \text{LogP}^2\rangle$ $- 0.688 \text{LogP}^3\rangle - 0.041 \text{LogP}^4\rangle + 0.027 \text{LogP}^6\rangle$	0.578	0.989	1.711	15.169	5.332	1.704	0.212	5.604
	$ Y_B^{\text{POL}}\rangle = -16485.827 1\rangle + 2491.049 \text{POL}\rangle - 146.094 \text{POL}^2\rangle$ $+ 4.037 \text{POL}^3\rangle - 0.047 \text{POL}^4\rangle + 2.9 \cdot 10^{-6} \text{POL}^6\rangle$	0.718	0.992	1.382	-0.355	-0.125	3.619	0.451	1.459
	$ Y_B^H\rangle = 4.876 1\rangle + 0.110 H\rangle + 0.004 H^2\rangle$ $- 2.3 \cdot 10^{-4} H^3\rangle - 7.67 \cdot 10^{-6} H^4\rangle + 6.3 \cdot 10^{-10} H^6\rangle$	0.856	0.996	1.163	19.088	6.709	9.349	1.166	6.908

^(a) the statistical Pearson correlation factor; ^(b) computed from Equation (7); ^(c) computed from Equation (9); ^(d) computed from Equation (10) with $t_{\text{Tabulated}}^{(0.99;20)} = 2.845$;

^(e) computed from Equation (11) with $F_{\text{Tabulated}}^{(0.99;1,21)} = 8.02$; ^(f) computed from Equation (8).

Table 6. Correlation equations for the Group-II models of Table 3 and the molecular structures and data of Table 4.

Catastrophe	QSAR Model	$R_{\text{Pearson}}^{(a)}$	$R_{\text{ALG}}^{(b)}$	$r^{(c)}$	$t\text{-Stud.}$	$t^{(d)}$	Fisher	$f^{(e)}$	$\Pi^{(f)}$
QSAR (II)	$ Y_{II}^{\text{LogP,POL}}\rangle = -2.044 1\rangle + 0.051 \text{LogP}\rangle + 0.242 \text{POL}\rangle$	0.556	0.989	1.778	-0.702	-0.245	4.464	0.763	1.9504
	$ Y_{II}^{\text{LogP,H}}\rangle = 5.379 1\rangle + 0.304 \text{LogP}\rangle - 0.023 H\rangle$	0.556	0.989	1.778	18.564	6.489	4.468	0.764	6.771
	$ Y_{II}^{\text{POL,H}}\rangle = -2.637 1\rangle + 0.248 \text{POL}\rangle - 0.021 H\rangle$	0.728	0.992	1.363	-1.151	-0.402	11.302	1.932	2.398

Table 6. Cont.

Catastrophe	QSAR Model	R _{Pearson} ^(a)	R _{ALG} ^(b)	r ^(c)	t- <i>Stud.</i>	t ^(d)	Fisher	f ^(e)	Π ^(f)
Hyperbolic umbilic (HU)	$\begin{aligned} Y_{HU}^{LogP,POL}\rangle &= -39.499 1\rangle - 2.463 LogP\rangle + 2.043 POL\rangle \\ &+ 0.104 (LogP)(POL)\rangle - 0.145 LogP^3\rangle - 6 \cdot 10^{-4} POL^3\rangle \end{aligned}$	0.715	0.992	1.387	-2.215	-0.774	3.561	0.609	1.701
	$\begin{aligned} Y_{HU}^{LogP,H}\rangle &= 5.319 1\rangle + 1.083 LogP\rangle - 0.002 H\rangle \\ &- 0.003 (LogP)(H)\rangle - 0.161 LogP^3\rangle - 9 \cdot 10^{-6} H^3\rangle \end{aligned}$	0.736	0.992	1.3485	19.328	6.756	4.019	0.687	6.923
	$\begin{aligned} Y_{HU}^{POL,H}\rangle &= -13.192 1\rangle + 0.766 POL\rangle + 0.122 H\rangle \\ &- 0.004 (POL)(H)\rangle - 2 \cdot 10^{-4} POL^3\rangle - 5.1 \cdot 10^{-7} H^3\rangle \end{aligned}$	0.755	0.993	1.315	-0.79	-0.276	4.503	0.770	1.549
Elliptic umbilic (EU)	$\begin{aligned} Y_{EU}^{LogP,POL}\rangle_A &= -69.262 1\rangle - 0.556 LogP\rangle + 4.531 POL\rangle \\ &+ 0.443 LogP^2\rangle - 0.068 POL^2\rangle \\ &+ 0.002 (LogP)(POL^2)\rangle - 0.322 LogP^3\rangle \end{aligned}$	0.757	0.993	1.312	-2.548	-0.891	3.582	0.612	1.670
	$\begin{aligned} Y_{EU}^{LogP,POL}\rangle_B &= 644.623 1\rangle + 0.022 LogP\rangle - 59.934 POL\rangle \\ &+ 0.467 LogP^2\rangle + 1.855 POL^2\rangle \\ &- 0.015 (POL)(LogP^2)\rangle - 0.019 POL^3\rangle \end{aligned}$	0.722	0.992	1.374	1.866	0.652	2.908	0.497	1.600
	$\begin{aligned} Y_{EU}^{LogP,H}\rangle_A &= 5.022 1\rangle + 0.974 LogP\rangle + 0.025 H\rangle \\ &+ 0.530 LogP^2\rangle + 0.001 H^2\rangle \\ &+ 2.87 \cdot 10^{-4} (LogP)(H^2)\rangle - 0.359 LogP^3\rangle \end{aligned}$	0.843	0.995	1.181	20.638	7.214	6.542	1.118	7.395

Table 6. Cont.

Catastrophe	QSAR Model	R _{Pearson} ^(a)	R _{ALG} ^(b)	r ^(c)	t- <i>Stud.</i>	t ^(d)	Fisher	f ^(e)	Π ^(f)
Elliptic umbilic (EU)	$\begin{aligned} Y_{EU}^{LogP,H}\rangle_B &= 4.779 1\rangle + 0.643 LogP\rangle + 0.029 H\rangle \\ &- 0.211 LogP^2\rangle + 0.004 H^2\rangle \\ &+ 0.001 (H)(LogP^2)\rangle + 5 \cdot 10^{-5} H^3\rangle \end{aligned}$	0.851	0.995	1.170	17.047	5.958	7.015	1.199	6.189
	$\begin{aligned} Y_{EU}^{POL,H}\rangle_A &= 807.822 1\rangle - 74.631 POL\rangle - 0.02 H\rangle \\ &+ 2.291 POL^2\rangle + 0.005 H^2\rangle \\ &- 2 \cdot 10^{-4} (POL)(H^2)\rangle - 0.023 POL^3\rangle \end{aligned}$	0.857	0.996	1.162	3.124	1.092	7.346	1.256	2.029
	$\begin{aligned} Y_{EU}^{POL,H}\rangle_B &= 11.888 1\rangle - 0.562 POL\rangle + 0.068 H\rangle \\ &+ 0.011 POL^2\rangle + 0.004 H^2\rangle \\ &- 4 \cdot 10^{-5} (H)(POL^2)\rangle + 4 \cdot 10^{-5} H^3\rangle \end{aligned}$	0.853	0.996	1.167	0.532	0.186	7.120	1.217	1.696
Parabolic umbilic (PU)	$\begin{aligned} Y_{PU}^{LogP,POL}\rangle_A &= 474.915 1\rangle + 0.021 LogP\rangle - 39.256 POL\rangle \\ &+ 0.454 LogP^2\rangle + 0.914 POL^2\rangle \\ &- 0.015 (LogP^2)(POL)\rangle - 10^{-4} POL^4\rangle \end{aligned}$	0.722	0.992	1.374	1.817	0.635	2.905	0.497	1.593
	$\begin{aligned} Y_{PU}^{LogP,POL}\rangle_B &= -67.522 1\rangle - 1.539 LogP\rangle + 4.444 POL\rangle \\ &+ 0.573 LogP^2\rangle - 0.067 POL^2\rangle \\ &+ 0.002 (POL^2)(LogP)\rangle - 0.115 LogP^4\rangle \end{aligned}$	0.703	0.992	1.411	-2.219	-0.776	2.611	0.446	1.671

Table 6. Cont.

Catastrophe	QSAR Model	R _{Pearson} ^(a)	R _{ALG} ^(b)	r ^(c)	t- <i>Stud.</i>	t ^(d)	Fisher	f ^(e)	Π ^(f)
Parabolic umbilic (PU)	$\begin{aligned} Y_{PU}^{LogP,H}\rangle_A &= 4.852 1\rangle + 0.700 LogP\rangle + 0.041 H\rangle \\ &- 0.240 LogP^2\rangle + 0.004 H^2\rangle \\ &+ 0.002 (LogP^2)(H)\rangle - 10^{-6} H^4\rangle \end{aligned}$	0.874	0.996	1.140	20.243	7.075	8.645	1.478	7.317
	$\begin{aligned} Y_{PU}^{LogP,H}\rangle_B &= 5.10 1\rangle + 0.552 LogP\rangle + 0.020 H\rangle \\ &+ 0.460 LogP^2\rangle + 9.57 \cdot 10^{-4} H^2\rangle \\ &+ 1.93 \cdot 10^{-4} (H^2)(LogP)\rangle - 0.099 LogP^4\rangle \end{aligned}$	0.767	0.993	1.295	16.828	5.882	3.815	0.652	6.058
	$\begin{aligned} Y_{PU}^{POL,H}\rangle_A &= 8.876 1\rangle - 0.366 POL\rangle + 0.069 H\rangle \\ &+ 0.008 POL^2\rangle + 0.003 H^2\rangle \\ &- 3.7 \cdot 10^{-5} (POL^2)(H)\rangle - 4.5 \cdot 10^{-7} H^4\rangle \end{aligned}$	0.841	0.995	1.183	0.386	0.135	6.447	1.102	1.623
	$\begin{aligned} Y_{PU}^{POL,H}\rangle_B &= 595.212 1\rangle - 48.906 POL\rangle - 0.019 H\rangle \\ &+ 1.129 POL^2\rangle + 5 \cdot 10^{-3} H^2\rangle \\ &- 1.49 \cdot 10^{-4} (H^2)(POL)\rangle - 1.73 \cdot 10^{-4} POL^4\rangle \end{aligned}$	0.856	0.996	1.163	3.074	1.074	7.292	1.246	2.015

^(a) the statistical Pearson correlation factor; ^(b) computed from Equation (7); ^(c) computed from Equation (9); ^(d) computed from Equation (10) with $t_{Tabulated}^{(0.99;19)} = 2.861$;

^(e) computed from Equation (11) with $F_{Tabulated}^{(0.99;2,20)} = 5.85$; ^(f) computed from Equation (8).

Table 7. Single-structure matrices of the Euclidean distances $\Delta\Pi_I$ of the QSAR and catastrophe models' relative statistics of Table 5 employing Equation (12).

Log P	F	C	ST	B
QSAR	1.750	2.645	2.905	3.627
F		2.411	1.732	2.865
C			1.437	1.174
ST				1.231

POL	F	C	ST	B
QSAR	0.517	1.198	0.828	0.830
F		1.317	0.717	0.524
C			0.670	0.983
ST				0.314

H	F	C	ST	B
QSAR	0.431	0.89	1.916	1.127
F		1.054	1.793	1.242
C			1.509	0.292
ST				1.29

Table 8. Differences $\Delta^2\Pi_I$ between the single-structure matrices of the Euclidean distances in Table 7.

Log P ÷ POL	F	C	ST	B
QSAR	1.233	1.446	2.076	2.797
F		1.094	1.015	2.341
C			0.767	0.191
ST				0.917

Log P ÷ H	F	C	ST	B
QSAR	1.32	1.755	0.988	2.501
F		1.358	0.062	1.624
C			0.072	0.882
ST				0.059

POL ÷ H	F	C	ST	B
QSAR	0.086	0.309	1.088	0.297
F		0.264	1.076	0.717
C			0.839	0.691
ST				0.976

Table 9. Single-structure matrices of the Euclidean distances $\Delta\Pi_{II}$ of the QSAR and catastrophe models' relative statistics of Table 6 employing Equation (12); note that for the degenerate models of Table 6 that one is employed that displays higher relative statistical power (Π).

Log P^POL	HU	EU	PU
QSAR	0.675	0.810	1.005
HU		0.139	1.414
EU			1.531

Log P^H	HU	EU	PU
QSAR	0.512	0.917	1.123
HU		0.964	0.878
EU			1.152

POL^H	HU	EU	PU
QSAR	1.170	1.652	1.640
HU		1.46	1.440
EU			0.02

For the trial set of molecules from Figure 1 and Table 4, the results in Tables 5 and 6 can be interpreted as follows:

- First, it is clear that consideration of the catastrophe (polynomial) correlations is an improvement over the old multi-linear QSAR statistics (see also Appendix-A2).
- The hydrophobicity indicator gives generally low correlations with any polynomial (linear, multilinear or catastrophe) approach, being a quite irrelevant linear QSAR descriptor (Table 5) but improving up to twice its influence within the swallow tail and butterfly phenomenologies once its fifth and sixth power involvement are considered. Nevertheless, this provides a sign of the value of catastrophe-QSAR for achieving a deeper understanding of the molecular mechanics of specific interactions when the normal multi-linear QSAR does not assign transport descriptors with much predictive power.
- The relative statistical power, as defined by Equation (8), does not always parallel the Pearson coefficient or the relative correlation factors, as is evident from Tables 5 and 6. However, because it includes more statistical information, we consider a model as relevant when it has greater individual output of this newly introduced statistical index. In particular, neither the linear nor the multilinear QSAR framework provides a good fit between the statistical correlation and the relative statistical power using the structural parameter combinations considered. Instead, parabolic catastrophe correlations, the *cusp and butterfly models*, are revealed to be quite relevant, in particular regarding the formation energy (H) for which they show the highest Pearson correlation and relative statistical power values in comparison with the other descriptors plugged into these models. Unfortunately, for the two-variable descriptor models of Table 6, no consistency was found between the highest Pearson value and the relative statistical power apart from a few degenerate cases of descriptors for the parabolic models where the highest relative statistical power value corresponds with the highest Pearson correlation. Note that for the degenerate cases of Table 6, when two mixed descriptors can be combined in two distinct ways, the working model is considered to have maximum relative statistical power.

However, because the two-fold aim of the present research is to find the best predictive model and the molecular mechanism of action for the given set of molecules, the statistical indices of Tables 5 and 6 are employed to compute the first- and second-order differences (or distances) in relative statistical power as described by Equations (12–15) of Section 3. They correspond to the inter-descriptor/inter-modeling paths of molecular actions, whose minimum values are identified according to the prescription of Equation (16).

Through this minimal relative statistical power path recipe, once the models and descriptors predicted to be on the forefront of the structure-action interaction are selected, they are then further filtered with the testing set to finally identify the best predictive model and reveal the mechanism of action by means of the structural descriptors considered.

In the present case of the HIV inhibitors in Table 4, the data computed from Tables 5 and 6 provide the results for Tables 7–9, to be discussed herein:

- Table 7: At the individual descriptor level, the cusp and butterfly models are very close to each other for Log P and the forming energy H, which is even more relevant for the hydrophobicity,

because for the forming energy it transpires from Table 5 that the butterfly model practically reduces to the cusp model because the sixth contribution virtually vanishes. However, for the structural influence on polarizability (POL) the butterfly and swallow tail are the closest models. When one considers the hierarchy of the individual descriptors according to their QSAR-I models in Table 5 in terms of the reduction in relative statistical power

$$\text{Log } P \rightarrow H \rightarrow \text{POL} \quad (17a)$$

through combining it with the catastrophes involved in Table 7, one correspondingly obtains the evolution cycle of the models:

$$(\dots \rightarrow) [Butterfly] \rightarrow [Cusp] \rightarrow [Butterfly] \rightarrow [SwallowTail] (\rightarrow \dots) \quad (17b)$$

- Table 8: When the second order distance difference is considered between the individual inter-modeling paths of Table 7, it can nevertheless be considered through the further variations of paths of Table 7. Also, the QSAR-I and the fold (F) catastrophe model intervene in changing the influence on specific interactions from POL to H. Therefore, by counting the minimum hierarchy of these paths, the distance ordering is obtained as follows:

$$(\text{Log } P \div H) \rightarrow (H \div \text{POL}) \rightarrow (\text{POL} \div \text{Log } P) \quad (18a)$$

which, remarkably, confirms the descriptors' cycles of influence in accordance with the first order prescription of Equation (17a). However, a more detailed succession is recorded for the inter-model evolution:

$$(\dots \rightarrow) [Butterfly] \cong [SwallowTail] \rightarrow [QSAR - I] \cong [Fold] \rightarrow [Cusp] \cong [Butterfly] (\dots \rightarrow) \quad (18b)$$

When comparing cycles (18b) with (17b), it seems that the QSAR-I and Fold models appear in (18b) at the second cycle after the first one is performed on the prescription of (17b). For this reason also, the direct second order inter-descriptor-inter-models analysis is undertaken, and the results are reported in Table 9, to be discussed hereafter.

- Table 9: Interestingly, in terms of the two structural descriptors, the QSAR model is present even though its individual statistics are not the highest in Table 6; however, judging by the ordering of minimum paths recorded, the coupling descriptors hierarchy is established as:

$$(H \& \text{POL}) \rightarrow (\text{POL} \& \text{Log } P) \rightarrow (\text{Log } P \& H) \quad (19a)$$

which is associated with the models' evolution

$$(\dots \rightarrow) [PU] \rightarrow [EU] \rightarrow [HU] \rightarrow [QSAR] (\rightarrow [PU] \dots \rightarrow) \quad (19b)$$

One should make "contact" between the descriptor hierarchies [(17a), (18a), (19a)] and the models' cycles [(17b), (18b) and (19b)] by means of the predictivity powers of the models along the minimum paths identified in Tables 7 and 9 with the single and double descriptors, respectively, for the non-Gaussian (NG) molecules of Table 4 and Figure 1. The results are systematically presented in Tables 10 and 11.

Table 10. Predicted activity as computed for the non-Gaussian molecules of Table 4 with the models of Table 5 founded along the minimum paths of Table 7; for each predicted model, its correlation with the observed activity is indicated at the bottom of the table.

Model Molecule	$ Y_C^{LogP}\rangle$	$ Y_C^H\rangle$	$ Y_{ST}^{POL}\rangle$	$ Y_B^{LogP}\rangle$	$ Y_B^{POL}\rangle$	$ Y_B^H\rangle$
NG1	5.586	6.179	5.294	5.094	-20.595	5.687
NG2	5.729	4.885	4.294	5.719	-9.764	4.360
NG3	5.676	0.415	4.708	5.531	-13.457	-7.932
NG4	5.729	6.156	5.149	6.657	-29.709	5.259
NG5	6.487	6.141	5.309	6.705	-25.700	5.923
NG6	6.399	5.438	5.258	6.708	-27.365	5.219
NG7	6.903	5.631	5.319	5.311	-21.540	5.984
NG8	6.904	5.334	5.027	5.995	-31.693	5.566
NG9	5.580	4.9357	5.328	5.054	-24.666	4.383
R-Pearson	0.195	0.129	0.174	0.701	0.488	0.026

Table 11. Predicted activity as computed for the non-Gaussian molecules of Table 4 with the models of Table 6 founded along the minimum paths of Table 9; for each predicted model, its correlation with the observed activity is indicated at the bottom of the table.

Model Molecule	$ Y_{II}^{LogP,H}\rangle$	$ Y_{HU}^{LogP,POL}\rangle$	$ Y_{HU}^{LogP,H}\rangle$	$ Y_{EU}^{LogP,POL}\rangle_A$	$ Y_{EU}^{POL,H}\rangle_A$	$ Y_{PU}^{POL,H}\rangle_B$
NG1	6.0865	5.918	5.308	5.387	5.351	7.210
NG2	5.581	5.839	5.399	5.448	4.816	4.578
NG3	6.785	6.132	7.526	5.686	1.423	7.234
NG4	7.115	6.642	6.037	6.289	5.480	7.765
NG5	6.495	7.382	6.853	7.277	6.033	7.629
NG6	6.163	7.291	6.426	7.104	7.338	7.647
NG7	5.790	7.388	6.087	7.615	6.879	6.547
NG8	5.761	7.560	6.330	7.640	7.895	7.447
NG9	5.467	5.755	4.786	5.177	7.586	7.303
R-Pearson	0.778	0.468	0.454	0.431	0.057	0.451

The results of correlation tests in Table 10 indicate the structure index–model activity hierarchy:

$$|Y_B^{LogP}\rangle > |Y_B^{POL}\rangle > |Y_C^{LogP}\rangle > |Y_{ST}^{POL}\rangle > |Y_C^H\rangle > |Y_B^H\rangle \quad (20)$$

Somehow the influences of POL and H are reversed relative to the prescription by trial succession of Equation (17a), revealing hydrophobicity as the main influential factor. However, due to the fact that the predicted activities of POL in Table 10 are all in the “opposite evolution direction” with respect to the activities recorded in Table 4, *i.e.*, they are all negative, the uni-parametric tests and their associated hierarchy (20) are discarded, and one looks toward the second class of QSAR and catastrophe algorithms.

Instead, the test correlations of Table 11 provide the structure-activity ordering for the bi-parameter-models

$$\left|Y_{II}^{LogP,H}\right\rangle > \left|Y_{HU}^{LogP,POL}\right\rangle > \left|Y_{HU}^{LogP,H}\right\rangle > \left|Y_{PU}^{POL,H}\right\rangle_B > \left|Y_{EU}^{LogP,POL}\right\rangle_A > \left|Y_{EU}^{POL,H}\right\rangle_A \quad (21)$$

Remarkably, the hierarchy (21) starts with the QSAR model, which is revealed to be at the top of the validated catastrophe models with statistical performance even higher than through the predicted equation of Table 6 and the trial set of Table 4. Moreover, the QSAR-II model involves parameters (Log P & H) that are followed by the hyperbolic umbilic (HU) model in terms of (Log P & POL) parameters, in this way recovering the original mono-structural influences as anticipated by Equations (17a) and (18a). Thus, the series of models in Equation (21) is validated, and it will be further employed to establish the models' successions and the molecular structural pattern of inhibiting anti-HIV-1 drug resistance. To this end, apart from the first and last models of Equation (21), which are associated with the maximum (0.778) and minimum (0.057) test performance, the middle catastrophe models provide closely related performance in the range (0.431, 0.468). Their graphical 3D-representation of the parametric domains Log P: (-1.50, 2.72), POL: (27.87, 38.48) and H: (-63.299, 17.808) of all (trial and test) structures in Table 4 are displayed in Figure 2. Next, it is apparent that they can be coupled according to the same spanned domains, thus forming the activity models' differences $\left|Y_{II}^{LogP,H}\right\rangle - \left|Y_{HU}^{LogP,H}\right\rangle$, $\left|Y_{HU}^{LogP,POL}\right\rangle - \left|Y_{EU}^{LogP,POL}\right\rangle_A$, $\left|Y_{EU}^{POL,H}\right\rangle_A - \left|Y_{PU}^{POL,H}\right\rangle_B$, plotted in the top of Figure 3. Through registering the parameters and the models' successions:

$$[QSAR] \xrightarrow{LogP,H} [HU] \xrightarrow{LogP,POL} [EU] \xrightarrow{POL,H} [PU] \quad (22)$$

one may reach the following important conceptual-computational conclusions:

- The HIV-1 inhibitory activity is triggered by a hydrophobic interaction followed by energetic stabilization of the ligand/substrate (pyridinone derivative/viral protein) interaction here modeled by the heat of molecular formation and eventually completed by the ionic field influence herein represented by the polarizability descriptor.
- Although the QSAR multi-linear model should not be excluded from the molecular modeling of complex bio-chemical interactions, it should be complemented with other polynomial correlational catastrophe-type models that produce significant results comparable to those of other 3D-modeling procedures such as docking-based comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) [24].

However, the issue remains of establishing the molecular structure most suitable for HIV-1 inhibitory activity among the considered pool of pyridinone derivatives in Table 4. To this end, the representations in Figure 3 are synergistically employed to identify the molecular structural domains that optimally promote binding of the pyridine derivative to the hydrophobic pocket in the p66 subunit of HIV-1 through searching for joint fulfillment of the following structural parameters and inter-model evolutionary generic principles:

- Log P: For positive values, the compound behaves hydrophobically and requires dissolution in an organic solvent; by contrast, for negative values the compound is hydrophilic and can be dissolved directly in an aqueous buffer. For Log P equal to 0, the compound partitions at a 1:1 organic-to-aqueous phase ratio, meaning that it is likely soluble in both organic and aqueous solvents and in cellular environments; thus, values of Log P equal to or greater than zero are

selected to achieve hydrophobicity and suitability for the cellular environment [43,44], while characterizing the stacking bonding of aromatic rings [45];

- H: Because the formation of a compound from its elements usually is an exothermic process, most heats of formation are negative, and this is also a characteristic of the dynamic equilibrium of ligand-substrate interactions [46]; note that the advantage of using heat of formation as QSAR descriptor resides in the following: it thermodynamically relates with the free energy $\Delta G = -RT \ln K_{eq}$ by the equilibrium constant K_{eq} which parallels the recorded activity at thermodynamic level [24]; it nevertheless expands the Gibbs free energy from the hydrogen to covalent bonding strength [45];
- PO: It is expected that “the natural direction of evolution of any system is towards a state of minimum polarizability” [47], while accounting for the dipolar interaction [45];
- Activity Models: Represent the same chemical-biological process providing their differences with respect to structural domains are minimized to zero.

Figure 2. 3D-representations of the QSAR and catastrophe activities for the tested models of Table 11 in the range of the structural indicators (Log P, Pol, H) as abstracted from Table 4.

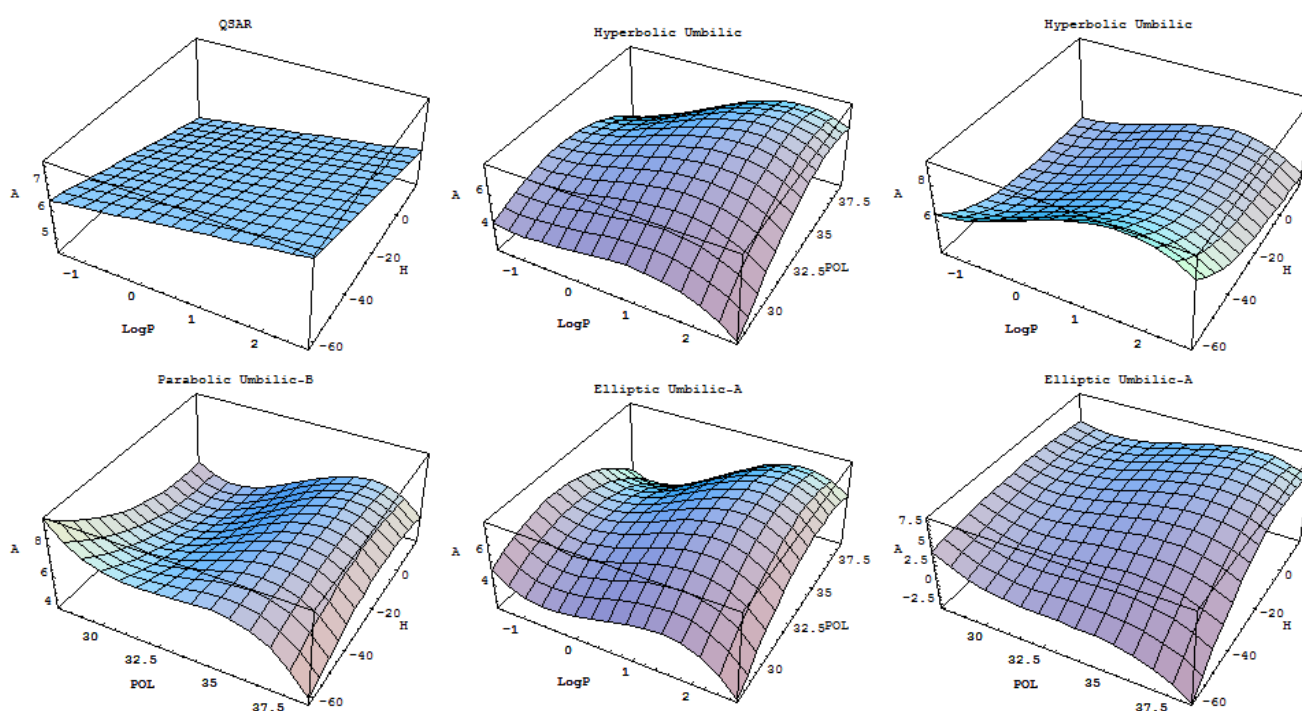
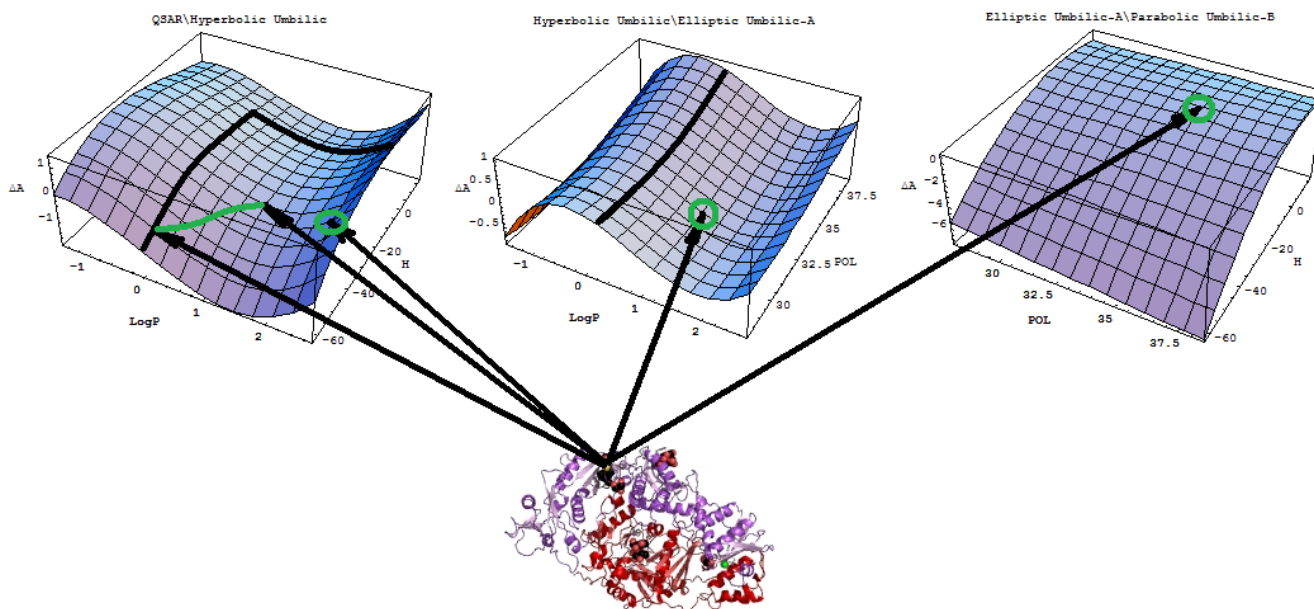


Figure 3. Determination of the structural domains of pyridinone-derivative type non-nucleoside reverse transcriptase inhibitors in the same range of structural descriptors by employing the principles of hydrophobicity, minimum polarizability, binding energy, and the minimum difference between the polynomial activity models of Figure 2; the hydrophobic pocket was identified in the p66 subunit of HIV-1-rt of specific transferase R221239 [48,49].



These principles are applied to the activity models' differences at the top of Figure 3, and they lead to the identification of the structural domain (and even points) characteristic of the pyridinone derivative most well-adapted to inhibiting the HIV-1 life cycle. The graphical results in Figure 3 suggest that the ordering of the structural indicators is:

$$\left| Y_{II}^{LogP,H} \right\rangle - \left| Y_{HU}^{LogP,H} \right\rangle : \{ \text{Log P: } (0,1.5) \ \& \ \text{H: } (-55,-40) \text{ kcal/mol} \} \cup \{ \text{Log P} \approx 2.5 \ \& \ \text{H} \approx -40 \text{ kcal/mol} \} \quad (23a)$$

$$\left| Y_{HU}^{LogP,POL} \right\rangle - \left| Y_{EU}^{LogP,POL} \right\rangle_A : \text{Log P} \approx 1 \ \& \ \text{POL} \approx 32 \text{ \AA}^3 \quad (23b)$$

$$\left| Y_{EU}^{POL,H} \right\rangle_A - \left| Y_{PU}^{POL,H} \right\rangle_B : \text{POL} \approx 34.5 \text{ \AA}^3; \ \text{H} \approx -10 \text{ kcal/mol} \quad (23c)$$

The “solution” of system (23) gives the actual molecules in Table 4 predicted to be the most potent binding inhibitors, namely compounds **27** (Log P \approx 2.72, H \approx -39.459 kcal/mol, POL \approx 35.55 \AA^3), **28** (Log P \approx 1.06, H \approx -34.478 kcal/mol, POL \approx 34.88 \AA^3), and **29** (Log P \approx 0.96, H \approx -21.361 kcal/mol, POL \approx 35.17 \AA^3). Most impressively, these molecules were also predicted by the much more sophisticated methods of CoMFA and CoMSIA as having increased binding affinity between the aromatic ring (or wing 2 of the pyridinone derivative) and amino acid Tyr181 of the first molecule and Tyr188 of the last two. These two amino acids are very important in the inhibition of RT by NNRTIs because the most common mutations are Tyr181Cys and Tyr188Cys, and they are responsible for the emergence of viruses resistant to pyridinone derivatives. Therefore, designing

pyridinone compounds that allow aromatic ring stacking interactions with Tyr181 and Tyr 188 may prevent these mutations and increase the activity of these anti-HIV drugs.

Overall, the QSAR presented here combined with catastrophe polynomial structure activity relationships provides a reliable conceptual and computational tool for identifying the mechanisms underlying ligand-substrate interactions and the structural domains best able to promote them. Consequently, this method should be further integrated into automated data processing and tested on other complex open systems with bio- or eco-toxicological relevance, especially where evolutionary life-cycles are present.

5. Conclusions

One of the most challenging battlefields in metabolic virology focuses on the complete and sustained inhibition of the HIV life cycle at its various levels. Thus: “an ideal anti-HIV agent should stop the virus’ progress and also the infection of healthy host cells, with no toxicity against normal cell physiology” [50]. Moreover, the ideal anti-HIV agent should avoid the drug-resistance phenomenon of HIV mutant variants. QSAR techniques are cost-effective computer-assisted drug design methods that can be used to obtain potential anti-HIV compounds with powerful biological effects and the lowest possible levels of side-effects and toxicity.

As the predictive roles of modeling and quantitative-structure-activity relationships (QSAR) in medicinal chemistry and drug synthesis are now recognized [51,52], thereby corroborating recent intriguing reports on the modest performance of direct statistical multilinear correlations in genotoxic carcinogenesis modeling of covalent drug binding to DNA followed by mutagenesis [53], the present study advances the idea of non-linear polynomial fits of observed/experimentally available $Activity = f(X_1, X_2)$, with X_1, X_2 being structural physicochemical parameters (usually hydrophobicity, polarizability and/or forming heat energy in accordance with the basic recommendation of Hansch) [54] under the seven polynomial forms inspired by Thom’s catastrophe theory [1] (see Table 3).

As an application of the emerging *catastrophe-QSAR* analysis to a recently reported set of pyridinone derivatives with non-nucleoside reverse transcriptase inhibitor activity, [24] all the modeling stages required by the OECD-QSAR principles [32] are implemented here in a synergistic manner, namely:

- (i) *A defined endpoint*: The hydrophobic binding of the inhibitor in the pocket of the p66 subunit of reverse-transcriptase was confirmed herein through the identification of hydrophobicity as the major influence among all the mono-nonlinear catastrophes employed; see Equation (17).
- (ii) *An unambiguous algorithm*: The Spectral-SAR minimum path principle [31,55–57] is here generalized to include relevant combination of statistical information (e.g., the correlation factor R , Student’s t -test, Fischer’s F -test) to provide an equal footing multi-dimensional Euler distance [see Equations (8–16)], thus avoiding the previously identified discrepancy in judging the mid-range performance in terms of correlation or other statistical factors [56].
- (iii) *A defined domain of applicability*: By performing linear vs. non-linear QSARs, the present strategy allows for the identification of recommended applicable structural domains through setting their difference to zero via inter-model activity minimization, which is equivalent to

assuring the “smoothness” of the inhibitor-protein binding evolution towards the final steric inhibition output.

- (iv) *Appropriate measures of goodness-of-fit, robustness and predictivity*: The trial results were evaluated by external validation employing a testing set, which was selected by means of Gaussian vs. non-Gaussian distributions of the compounds’ activities, an improvement over the earlier arbitrariness of sampling the compounds only within a certain activity range. For instance, for linear QSAR the predicted correlation was superior to the tested correlation, thus confirming the reliability of this validation technique.
- (v) *A mechanistic interpretation*: The selected succession of catastrophe-QSARs indicates that the inhibitor-HIV protein binding mutations that are involved in “birth and death” processes are associated with “waves” of induced activity in certain structural domain variants (see Figure 2). Moreover, the flat QSAR hypersurface should be complemented with catastrophe analysis to determine the specific structural domains for optimum interactions (see Figure 3) and for the associated molecular structure design of NNRT inhibitors.

Because the catastrophe-QSAR approach was found to successfully identify the molecular compounds with the most anti-HIV-1 potency as predicted by other 3D-QSAR methods, these results encourage further applications and implementations of Thom’s non-linear correlations with the goal of analytically modeling complex dynamic ligand-receptor interactions, especially on the molecular fragment or structural alert level [41], on a chemometric basis.

Acknowledgements

This work was supported by CNCS-UEFISCDI (former CNCSIS-UEFISCSU) project TE16/2010-2011 within the PN II-RU-TE-2009-1 framework, and partly by the grant POSDRU/21/1.5/G/38347.

References

1. Thom, R. *Stabilité Structurelle et Morphogénèse*; Benjamin-Addison-Wesley: New York, NY, USA, 1973.
2. Viret, J. Reaction of the organism to stress: The survival attractor concept. *Acta Biotheor.* **1994**, *42*, 99–109.
3. Lacorre, P. Predation and generation processes through a new representation of the cusp catastrophe. *Acta Biotheor.* **1997**, *45*, 93–115.
4. Viret, J. Topological approach of Jungian psychology. *Acta Biotheor.* **2010**, *58*, 233–245.
5. Cerf, R. Catastrophe theory enables moves to be detected towards and away from self-organization: The example of epileptic seizure onset. *Biol. Cybern.* **2006**, *94*, 459–468.
6. Silvi, B.; Savin, A. Classification of chemical bonds based on topological analysis of electron localization functions. *Nature* **1994**, *371*, 683–686.
7. Putz, M.V. Markovian approach of the electron localization functions. *Int. J. Quantum Chem.* **2005**, *105*, 1–11.

8. Aerts, D.; Czachor, M.; Gabora, L.; Kuna, M.; Posiewnik, A.; Pykacz, J.; Syty, M. Quantum morphogenesis: A variation on Thom's catastrophe theory. *Phys. Rev.* **2003**, *67*, doi:10.1103/PhysRevE.67.051926.
9. De Clercq, E. Anti-HIV drugs: 25 compounds approved within 25 years after the discovery of HIV. *Int. J. Antimicrob. Agents* **2009**, *33*, 307–320.
10. De Clercq, E. The history of antiretrovirals: Key discoveries over the past 25 years. *Rev. Med. Virol.* **2009**, *19*, 287–299.
11. El Safadi, Y.; Vivet-Boudou, V.; Marquet, R. HIV-1 reverse transcriptase inhibitors. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 723–737.
12. Ivetac, A.; McCammon, J.A. Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multi-copy molecular dynamics simulations. *J. Mol. Biol.* **2009**, *388*, 644–658.
13. Gupta, S.P. Advances in QSAR studies of HIV-1 reverse transcriptase inhibitors. *Prog. Drug Res.* **2002**, *58*, 223–264.
14. 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.
15. Prajapati, D.G.; Ramajayam, R.; Yadav, M.R.; Giridhar, R. The search for potent, small molecule NNRTIs: A review. *Bioorg Med. Chem.* **2009**, *17*, 5744–5762.
16. Zhan, P.; Chen, X.; Li, D.; Fang, Z.; de Clercq, E.; Liu, X. HIV-1 NNRTIs: Structural diversity, pharmacophore similarity, and implications for drug design. *Med. Res. Rev.* **2011**, in press.
17. Chen, X.; Zhan, P.; Li, D.; De Clercq, E.; Liu, X. Recent advances in DAPYs and related analogues as HIV-1 NNRTIs. *Curr. Med. Chem.* **2011**, *18*, 359–376.
18. Rebehmed, J.; Barbault, F.; Teixeira, C.; Maurel, F. 2D and 3D QSAR studies of diarylpyrimidine HIV-1 reverse transcriptase inhibitors. *J. Comput. Aided Mol. Des.* **2008**, *22*, 831–841.
19. Afantitis, A.; Melagraki, G.; Sarimveis, H.; Koutentis, P.A.; Markopoulos, J.; Igglessi-Markopoulou, O. A novel simple QSAR model for the prediction of anti-HIV activity using multiple linear regression analysis. *Mol. Divers.* **2006**, *10*, 405–414.
20. Marino, D.J.G.; Castro, E.A.; Toropov, A. Improved QSAR modeling of anti-HIV-1 activities by means of the optimized correlation weights of local graphs invariants. *Central Eur. J. Chem.* **2006**, *4*, 135–148.
21. Mandal, A.S.; Roy, K. Predictive QSAR modeling of HIV reverse transcriptase inhibitor TIBO derivatives. *Eur. J. Med. Chem.* **2009**, *44*, 1509–1524.
22. Bak, A.; Polanski, J. A 4D-QSAR study on anti-HIV HEPT analogues. *Bioorg. Med. Chem.* **2006**, *14*, 273–279.
23. Duda-Seiman, C.; Duda-Seiman, D.; Dragoş, D.; Medeleanu, M.; Careja, V.; Putz, M.V.; Lacrămă, A.-M.; Chiriac, A.; Nuțiu, R.; Ciubotariu, D. Design of anti-HIV ligands by means of minimal topological difference (MTD) Method. *Int. J. Mol. Sci.* **2006**, *7*, 537–555.
24. Medina-Franco, J.L.; Rodríguez-Morales, S.; Juárez-Gordiano, C.; Hernández-Campos, A.; Castillo, R. Docking-based CoMFA and CoMSIA studies of non-nucleoside reverse transcriptase inhibitors of the pyridinone derivative type. *J. Comput. Aided Mol. Des.* **2004**, *18*, 345–360.
25. Topliss, J. *Quantitative Structure-Activity Relationships of Drugs*; Academic Press: New York, NY, USA, 1983.

26. Seyfel, J.K. *QSAR and Strategies in the Design of Bioactive Compounds*; VCH Weinheim: New York, NY, USA, 1985.
27. Duchowicz, P.R.; Castro, E.A. *The Order Theory in QSPR-QSAR Studies*; Mathematical Chemistry Monographs, University of Kragujevac: Kragujevac, Serbia, 2008.
28. Zhao, V.H.; Cronin, M.T.D.; Dearden, J.C. Quantitative structure-activity relationships of chemicals acting by non-polar narcosis—Theoretical considerations. *Quant. Struct. Act. Relat.* **1998**, *17*, 131–138.
29. Pavan, M.; Netzeva, T.; Worth, A.P. Review of literature based quantitative structure-activity relationship models for bioconcentration. *QSAR Comb. Sci.* **2008**, *27*, 21–31.
30. Pavan, M.; Worth, A.P. Review of estimation models for biodegradation. *QSAR Comb. Sci.* **2008**, *27*, 32–40.
31. Putz, M.V.; Putz, A.M. Timisoara Spectral—Structure Activity Relationship (Spectral-SAR) Algorithm: From statistical and algebraic fundamentals to quantum consequences. In *Quantum Frontiers of Atoms and Molecules*; Putz, M.V., Ed.; NOVA Science Publishers Inc.: New York, NY, USA, 2011; Chapter 21, pp. 539–580.
32. *OECD Principles: Guidance Document on the Validation of (Q)SAR Models*. OECD Environment Directorate: Paris, France, 2007.
33. Putz, M.V.; Putz, A.M.; Barou, R. Spectral-SAR Realization of OECD-QSAR Principles. *Int. J. Chem. Model.* **2011**, *3*, in press.
34. Krokidis, X.; Noury, S.; Silvi, B. Characterization of elementary chemical processes by catastrophe theory. *J. Phys. Chem. A* **1997**, *101*, 7277–7282.
35. Putz, M.V. Path integrals for electronic densities, reactivity indices, and localization functions in quantum systems. *Int. J. Mol. Sci.* **2009**, *10*, 4816–4940.
36. Weisstein, E.W. Catastrophe. From *MathWorld*—A Wolfram Web Resource. Available online: <http://mathworld.wolfram.com/Catastrophe.html> (accessed on 1 September 2011).
37. Sanns, W. *Catastrophe Theory with Mathematica: A Geometric Approach*; DAV: Waghäusel, Germany, 2000.
38. Lu, X.-F.; Chen, Z.-W. The development of anti-HIV-1 drugs. *Acta Pharm. Sin.* **2010**, *45*, 165–176.
39. Putz, M.V. Residual-QSAR. Implications for genotoxic carcinogenesis. *Chem. Central J.* **2011**, *5*, doi:10.1186/1752-153X-5-29.
40. Putz, M.V.; Lazea, M.; Sandjo, L.P. Quantitative Structure Inter-Activity Relationship (QSInAR). Cytotoxicity study of some hemisynthetic and isolated natural steroids and precursors on human fibrosarcoma cells HT1080. *Molecules* **2011**, *16*, 6603–6620.
41. Putz, M.V.; Ionașcu, C.; Putz, A.M.; Ostafe, V. Alert-QSAR. Implications for electrophilic theory of chemical carcinogenesis. *Int. J. Mol. Sci.* **2011**, *12*, 5098–5134.
42. Hypercube, Inc. *HyperChem 7.01 [Program Package]*; Hypercube, Inc.: Gainesville, FL, USA, 2002.
43. Leo, A.; Hansch, C.; Elkins, D. Partition coefficients and their uses. *Chem. Rev.* **1971**, *71*, 525–616.
44. Cronin, D.; Mark, T. The role of hydrophobicity in toxicity prediction. *Curr. Comput. Aided Drug Design* **2006**, *2*, 405–413.

45. Selassie, C.D. History of Quantitative Structure-Activity Relationships. In *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Abraham, D.J., Ed.; Wiley: New York, NY, USA, 2003; pp. 1–48.
46. Masterton, W.L.; Slowinski, E.J.; Stanitski, C.L. *Chemical Principles*; CBS College Publishing: Philadelphia, PA, USA, 1983.
47. Chattaraj, P.K.; Sengupta, S. Popular electronic structure principles in a dynamical context. *J. Phys. Chem.* **1996**, *100*, 16126–16130.
48. Himmel, D.M.; Das, K.; Clark, A.D.; Hughes, S.H.; Benjahad, A.; Oumouch, S.; Guillemont, J.; Coupa, S.; Poncelet, A.; Csoka, I.; *et al.* Crystal structures for HIV-1 reverse transcriptase in complexes with three pyridinone derivatives: A new class of non-nucleoside inhibitors effective against a broad range of drug-resistant strains. *J. Med. Chem.* **2005**, *48*, 7582–7591.
49. The European Bioinformatics Institute. Available online: <http://www.ebi.ac.uk/pdbsum/2BE2> (accessed on 11 September 2011).
50. Duda-Seiman, C.; Duda-Seiman, D.; Putz, M.V.; Ciubotariu, D. QSAR modeling of anti-HIV activity with HEPT derivatives. *Digest J. Nanomat. Biostruct.* **2007**, *2*, 207–219.
51. Croce, C.M. Oncogenes and cancer. *N. Engl. J. Med.* **2008**, *358*, 502–511.
52. Dingli, D.; Nowak, M.A. Cancer biology: Infectious tumour cells. *Nature* **2006**, *443*, 35–36.
53. Benigni, R.; Bossa, C.; Jeliaskova, N.; Netzeva, T.; Worth, A. *The Benigni/Bossa rules for mutagenicity and carcinogenicity—A module of Toxtree*; European Commission report EUR 23241; Office for Official Publications of the European Communities: Luxembourg, 2008; pp. 1–69.
54. Hansch, C.; Kurup, A.; Garg, R.; Gao, H. Chem-bioinformatics and QSAR: A review of QSAR lacking positive hydrophobic terms. *Chem. Rev.* **2001**, *101*, 619–672.
55. 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.
56. Putz, M.V.; Putz, A.M.; Lazea, M.; Chiriac, A. Spectral vs. statistic approach of structure-activity relationship. Application on ecotoxicity of aliphatic amines. *J. Theor. Comput. Chem.* **2009**, *8*, 1235–1251.
57. *QSAR & Spectral-SAR in Computational Ecotoxicology*; Putz, M.V., Ed.; Apple Academics: Ontario, Canada, 2012, in press.
58. Zeeman, E.C. Catastrophe theory. *Sci. Am.* **1976**, *234*, 65–83.
59. Morse, M. The critical points of a functional on n variables. *Trans. Am. Math. Soc.* **1931**, *33*, 72–91.
60. Arnold, V.I. Local normal forms of functions. *Invent. Math.* **1976**, *35*, 87–109.
61. Poston, T.; Stewart, I. *Catastrophe Theory and Its Applications*; Pitman Publishing: Boston, MA, USA, 1978.
62. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.

Appendix

A1. More on Catastrophe Theory Background

The foreground of the Catastrophe Theory lies on expressing the Taylor series associated to a smooth function $\eta(c, x)$, $(c, x) = (c_1, \dots, c_k, x_1, \dots, x_m)$, say in its origin $(c, x) = 0$ under the form

$$\eta(c, x) = j^s \eta(c, x) + \text{tayl} \tag{A.1}$$

viewed as the summation of the so called *s-jet* or *s-current*

$$j^s f(c, x) = \sum_{r=0}^s \frac{1}{r!} D^r \eta|_0 x^r \tag{A.2}$$

and of its tail generically called here as “*tayl*”. However, in modeling the natural phenomena, unlike the regular (like planets orbits) or continuous ones (with small perturbations included) where the truncation to the *s-jet* works fine, many of registered events display sudden (or “catastrophic”) characters, like earthquakes, population growth, or cancer spreading, thus highly requiring for counting of the Taylor tail as well; such need was elegantly resumed by C. E. Zeeman, one of the pioneers of Catastrophe Theory [58], by “allowing the tail of the Taylor series to wag the dog”. When the *tayl* part is becoming important it shapes as the quadratic type dependency on the control \times behavior joint space where the original function was defined:

$$\text{tayl} \propto g^2(c, x) \tag{A.3}$$

This is due to the celebrated Morse’s bifurcation lemma [59] around the so called critical points of the original function, see Equation (3) of the main text, where it actually equivalent the original function with the family of function

$$\eta(c, x) \rightarrow \tilde{\eta}(g_1(c, x), \dots, g_s(c, x)) \pm g_{s+1}^2 \pm \dots \pm g_m^2 \tag{A.4}$$

Here *s-stays* also the co-rank of the Hessian of $\eta(c, x)$ in the point $(c, x) = 0$. The main question that arises hereby is to try to identify the so called local types of function in a *k*-parametric (control space) family of functions, or, even more, being given a function to identify in its neighborhood the family it belongs to. The solution to this problem was furnished by Thom [1] and then by Arnold [60], by using the powerful concepts of co-dimension and structural transversality, such that the resulted classification theorem formulates the seven elementary so called catastrophe function of Table 2 as governing all the natural phenomena where the co-dimension is no greater than 4 (four). To better understand that this is indeed covering quite general plethora of natural dynamic systems (with complicated local/turning/singular points modeling sudden changes), enough recalling the heuristic example of the co-dimension for England-Scotland frontier, for instance, that is always equal to 1 (one) no matter one represents the frontier as a line (the road along it), as bidimensional (the road through it on the Earth), as tridimensional (the road through it by plain), or as 4-D (in relativistic vision when the space-time cone is considered as well along it) [61]. It is this co-dimension that controls so powerfully the reduction of all possible power expansions of smooth functions to those seven presented on Table 2; there, one sees the co-dimension number is always equal with the number of parameters from

the control space appearing in the Thom polynomials; they, in fact, represent families of functions, *i.e.*, controlling large classes of functions that drive open systems in similar (local) ways. In taxonomical (or algebraically) terms, it is said that although not all functions are typical (or elementary) their families are typical as families; In analytically terms, as all minima through origin look the same (there are said to be typical, and typical like the Morse minima of generalized parabola $x_1^2 + \dots + x_m^2$, eventually after re-parameterization) likewise any transverse path through any non-Morse function that can be found within a family of finite functions looks the same as all other transverse paths in the family (those of Table 2). Even more, the co-rank of those functions (as the co-rank of their Hessian on the critical/singular/turning points) fixes also the minimum of variables that function can be reduced to; for example, if a function of 2011 variables has a critical point of co-rank equal 1, the actual function to be studied is of only 1 variable! This makes the Catastrophe Theory extremely interesting for being applied on QSAR studies, where the available structural variables are listed on hundred pages [62], while in fact one searches for modeling functions that enter natural classes or family of functions with an universal character—as the Thom polynomials are—and therefore aiming to work with appropriate functions with considerable lower number of variables/structural descriptors, see Table 3.

A2. Catastrophe Theory Implication on Pearson Correlation

Since the transformation of the original smooth function into catastrophe one involves the Morse parabolic polynomials contribution, see Equation (A.4), one may employ this recipe to consider the ordinary QSAR predicted activity, say Y^{QSAR} , and of its transformation into the Catastrophe-QSAR one, say $Y^{\Gamma/QSAR}$, through the Gaussian mapping

$$Y_{i=1,N}^{\Gamma/QSAR}(X_1, \dots, X_M) = Y_{i=1,N}^{QSAR}(X_1, \dots, X_M) \pm Y_{i=1,N}^{M/\Gamma/QSAR}(X_2, \dots, X_M) \quad (\text{A.5})$$

with

$$Y_{i=1,N}^{M/\Gamma/QSAR}(X_2, \dots, X_M) = \sum_{j=2}^M \frac{1}{\sqrt{\pi}} \exp\left(-\frac{X_j^2}{4\sigma_{i=1,N}^2}\right) \quad (\text{A.6})$$

while referring to the running-indices assumed in Table 1. The form (A.5) with (A.6) recovers the original QSAR predicted function/value when all dispersions over all structural variables vanish

$$Y_{i=1,N}^{\Gamma/QSAR}(X_1, \dots, X_M) \xrightarrow{\sigma_{i=1,N} \rightarrow 0} Y_{i=1,N}^{QSAR}(X_1, \dots, X_M) \quad (\text{A.7})$$

thus motivating the actual generalization for treating the natural non-zero dispersive phenomena. On the other side, for higher dispersive values of structural variables (*i.e.*, when their domains of applicability eventually overlap and promote interactions, *i.e.*, the appearance of cross products in Tables 2 and 3) it produces the second order development

$$Y_{i=1,N}^{\Gamma/QSAR}(X_1, \dots, X_M) \xrightarrow{\sigma_{i=1,N} \rightarrow \infty} Y_{i=1,N}^{QSAR}(X_1, \dots, X_M) \pm \tilde{X}_2^2 \pm \dots \pm \tilde{X}_M^2 \quad (\text{A.8})$$

under appropriate transformations $\tilde{X}_{j=2,M} = \tilde{X}_{j=2,M}(X_{j=1,M})$. However, one can see that in the Catastrophe Theory's language the first function of the right hand side in (A.8) stays for the 1-jet for

the function Y^{QSAR} , while the hole expression (A.8) having the Hessian co-rank of order 2 is in full consistence with the maximum co-rank universal unfolding for the polynomials of Table 2.

Next, one likes to check for the effect the Gaussian development (aka the catastrophe transformation) of (A.5) has on the statistical (Pearson) statistical coefficient respecting the QSAR value

$$R_0 = \sqrt{1 - \frac{1}{\sigma_A} \sum_{i=1}^N (A_i - Y_i^{QSAR})^2}, \quad \sigma_A = \sum_{i=1}^N (A_i - \bar{A})^2 \quad (\text{A.9})$$

For the sake of clarity we will chose only one sign on (A.5), while the result will not depend on it, and successively obtain

$$\begin{aligned} R_\Gamma &= \sqrt{1 - \frac{1}{\sigma_A} \sum_{i=1}^N (A_i - Y_i^{\Gamma/QSAR})^2} \\ &= \sqrt{R_0 - \frac{1}{\sigma_A} \sum_{i=1}^N (Y_i^{M/\Gamma/QSAR})^2 + \frac{2}{\sigma_A} \sum_{i=1}^N (A_i - Y_i^{QSAR})(Y_i^{M/\Gamma/QSAR})} \\ &\leq \sqrt{R_0 - \frac{1}{\sigma_A} \sum_{i=1}^N (Y_i^{M/\Gamma/QSAR})^2} + 2 \sqrt{\frac{1}{\sigma_A} \sum_{i=1}^N (A_i - Y_i^{QSAR})^2} \sqrt{\frac{1}{\sigma_A} \sum_{i=1}^N (Y_i^{M/\Gamma/QSAR})^2} \end{aligned} \quad (\text{A.10})$$

where in the last relation the Cauchy-Schwarz inequality was used:

$$\sum_{i=1}^n u_i v_i \leq \sqrt{\sum_{i=1}^n u_i^2} \sqrt{\sum_{i=1}^n v_i^2} \quad (\text{A.11})$$

Next, in order to draw results that do not depend either on M -the number of structure variables nor on N -the number of chemicals/molecules involved in a custom QSAR study, one assumes dealing with the same dispersion of the observed activity as well as for each descriptor (the so called homogeneous assumption, $\sigma = \sigma_A = \sigma_i, \forall i = \overline{1, N}$) likely to be valid when dealing with great number of structural descriptors; this way, one actually performs the asymptotic limit $M \rightarrow \infty$ on (A.6) for all $i = \overline{1, N}$ and recognizes the Poisson integral result

$$Y^{\infty/\Gamma/QSAR} = \int_0^\infty \frac{1}{\sqrt{\pi}} \exp\left(-\frac{X^2}{4\sigma}\right) dX = \sqrt{\sigma} \quad (\text{A.12})$$

Accordingly, the inequality (A.10) now reads

$$R_{\infty/\Gamma}^2 \leq R_0^2 - N + 2\sqrt{N(1 - R_0^2)} \quad (\text{A.13})$$

It may be rearranged upon the second order equation in N -chemicals' space

$$-N + 2\sqrt{(1 - R_0^2)}\sqrt{N} + (R_0^2 - R_{\infty/\Gamma}^2) \geq 0 \quad (\text{A.14})$$

whose universal fulfillment leads with the condition

$$R_{\infty/\Gamma} \geq 1 \quad (\text{A.15})$$

Since the result (A.15) was obtained within asymptotic conditions regarding the number of structural descriptors and homogeneous dispersion against recorded activity, it can be naturally asserted to its minimum as

$$R_{M/T} \rightarrow 1 \geq \forall R_0 \quad (\text{A.16})$$

thus heuristically proving the superiority for the catastrophe-QSAR modeling over the fashioned QSAR, therefore further motivating the present approach. As numerical illustration of the general prescription of inequality (A.16) the present application confirms it by all one-to-one (*i.e.*, catastrophe-QSAR vs. simple QSAR) results reported in Tables 5 and 6.

© 2011 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).