J. Electrochem. Sci. Eng. **15(1)** (2025) 2374; http://dx.doi.org/10.5599/jese.2374

Open Access : : ISSN 1847-9286
www.jESE-online.org

Original scientific paper

Extended characteristic polynomial estimating the electrochemical behaviour of some 4-(azulen-1-yl)-2,6-divinylpyridine derivatives

Eleonora-Mihaela Ungureanu[1], Amalia Ștefaniu[2], Raluca Isopescu[3], Cornelia-Elena Mușina[3], Magdalena-Rodica Bujduveanu[3] and Lorentz Jäntschi[4,5,✉]

[1]*Doctoral School of Chemical Engineering and Biotechnologies, National University of Science and Technology POLITEHNICA, Bucharest, Romania*
[2]*National Institute of Chemical, Pharmaceutical Research and Development, Bucharest, Romania*
[3]*Faculty of Chemical Engineering and Biotechnologies, National University of Science and Technology POLITEHNICA, Bucharest, Romania*
[4]*Department of Physics and Chemistry, Technical University of Cluj-Napoca, Romania*
[5]*Laboratory of Electrochemistry For Advanced Materials, Technical University of Cluj-Napoca, Cluj-Napoca, Romania*

Corresponding authors: ✉ *lorentz.jantschi@chem.utcluj.ro; tel.: +4-0264-401-775*

Received: June 5, 2024; Accepted: July 18, 2024; Published: July 26, 2024

Abstract

Six derivatives of 4-(azulen-1-yl)-2,6-divinylpyridine were the subject of experimental determination of oxidation and reduction potentials being reported elsewhere. In this paper, a computational study was employed in order to obtain a function of structure for these potentials. The geometry was optimized at three theory levels (MMFF94, B3LYP and M06), and the following analysis was conducted with the separately saved optimum geometry in each instance. Two families of molecular descriptors (FMPI and EChP) were used to derive structure-based descriptors. Simple linear regressions were extracted with the best of descriptors for each family and level of theory for both potentials. The study revealed that the MMFF94 optimum geometries best explained the selected electrochemical properties. Furthermore, the EChP family of descriptors, much bigger than FMPI (about 64 times), was able to better explain the connection between the structure and the property. Once more, it has been shown that the eigenproblem has deep roots in structural chemistry.

Keywords

Oxidation potential; reduction potential; structure - activity relationship; eigenproblem

Introduction

Substituting pyridine (CAS RN 110-86-1) leads to a key six-membered heterocyclic scaffold (Figure 1), structurally related to benzene and having a conjugated system of six π electrons delocalized over the ring.

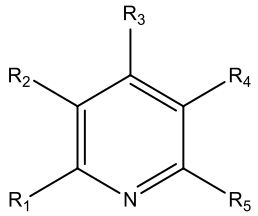
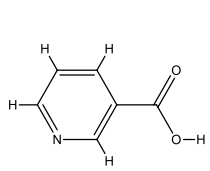
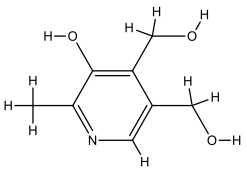


Figure 1. *Pyridine scaffold*

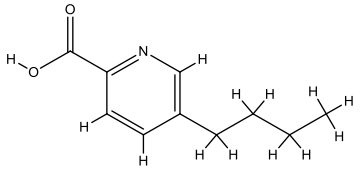
It occurs in natural compounds, drug molecules and vitamins, being a precursor of the synthesis of larger molecules [1]. Pyridine derivatives have multiple uses. Figure 2 depicts, in increasing complexity, some pyridine derivatives.



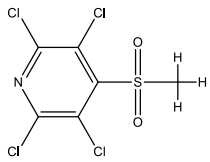
(a) CAS RN 59-67-6



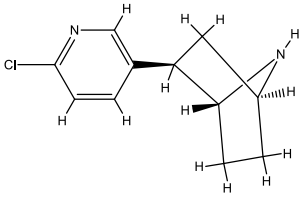
(b) CAS RN 65-23-6



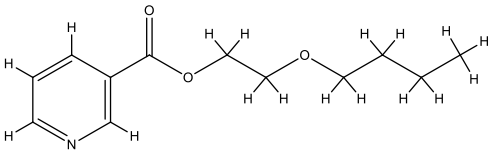
(c) CAS RN 536b-69-6



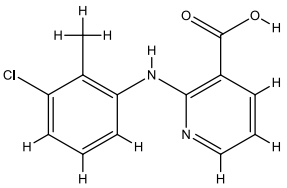
(d) CAS RN 13108-52-6



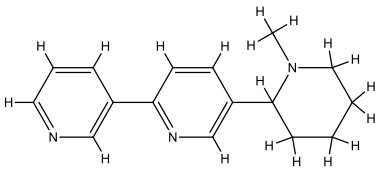
(e) CAS RN 140111-52-0



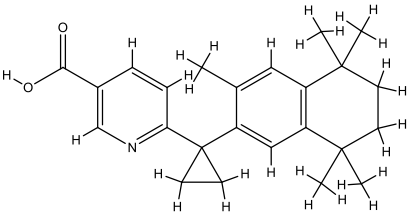
(f) CAS RN 13912-80-6



(g) CAS RN 17737-65-4



(h) CAS RN 20410-87-1



(i) CAS RN 153559-76-3

Figure 2. *Some pyridine derivatives*

Nicotinic acid (CID 938, $C_6H_5NO_2$, Figure 2a) is vitamin B3, an essential nutrient [2]; Pyridoxine (CID 1054, $C_8H_{11}NO_3$, Figure 2b) is vitamin B6, an important dietary supplement [3]; Fusaric acid (CID 3442, $C_{10}H_{13}NO_2$, Figure 2c) is a wilting agent and antibiotic [4]; Davicil (CID 61579, $C_6H_3Cl_4NO_2S$, Figure 2d) is a fungicide and allergen [5]; Epibatidine (CID 854023, $C_{11}H_{13}ClN_2$, Figure 2e) causes numbness and paralysis and may cause respiratory arrest [6]; Nicoboxil (CID 14866, $C_{12}H_{17}NO_3$, Figure 2f) is an analgesic effective on acute lower back pain [7]; Clonixin (CID 28718, $C_{13}H_{11}ClN_2O_2$, Figure 2g) is an anti-inflammatory, analgesic, antipyretic, platelet aggregation inhibitor [8]; Anabasamine (CID 161313, $C_{16}H_{19}N_3$, Figure 2h) is an inhibitor of acetylcholinesterase [9]; lastly, LG100268 (CID 3922, $C_{24}H_{29}NO_2$, Figure 2i) is a potent rexinoid [10].

As exemplified in Figure 2, the physical properties and biological activities of pyridine-containing compounds can be tuned by functional groups added to the scaffold. Of medical interest is the use of pyridine derivatives for making potent drugs. Table 1 contains some representative examples for medical use.

Table 1. *Some biomedical applications of pyridines derivatives*

Activity	Derivative(s)	Ref.
Antibacterial	Pyrazolo[3,4-b] Pyridine	[11]
Antibiotic	Streptonigrin	[12]
Antidiabetic	Glicaramide	[13]
Antiestrogenic	Tamoxifen analogue	[14]
Antifungal	Tetrahydroimidazo[1,2-a]pyridine	[15]
Antimicrobial	Imidazo[1,2-a]pyridine	[16]
Antimycobacterial	Isoniazid	[17]
Antineuralgic	Decumbenine B	[18]
Antioxidant	Imidazo[4,5-b]pyridines	[19]
Antiparasitic	Azamethiphos	[20]
Antispasmodic	Papaverine	[21]
Antitubercular	Ethionamide	[22]
Antitumor	Pyrrolo-pyridine benzamide	[23]
Antiviral	Indeno[1,2-b]pyridine	[24]

On the other hand, if the chemical change is made into solutions, then it usually has a difference of electrical potential associated, and electrochemical methods may provide further insights about it, such as the mechanism of redox reaction, but can also be a powerful tool for the synthesis of new compounds on a gram scale [25]. The mechanisms can be treacherous and may involve models with a great deal of mathematics and numerical solving incorporated [26].

Materials and methods

Vinyl-pyridines

The 4-(azulen-1-yl)-2,6-divinylpyridine scaffold (see Figure 3) has been used to obtain and characterize six derivatives (Table 2), where experimental oxidation (E_a) and reduction (E_c) potentials were taken from [27].

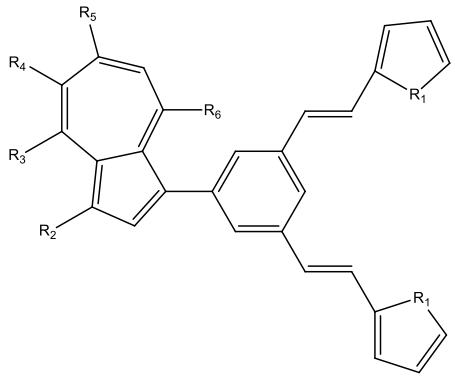


Figure 3. *The 4-(azulen-1-yl)-2,6-divinylpyridine template*

Table 2. *Templating of 4-(azulen-1-yl)-2,6-divinylpyridine derivatives*

Figure	R_1	R_2	R_3	R_4	R_5	R_6	Molecular formula	E_a / V	E_c / V
4a	O	CH_3	H	i-Pr	H	CH_3	$C_{32}H_{29}NO_2$	0.318	−2.071
4b	O	H	CH_3	H	CH_3	CH_3	$C_{30}H_{25}NO_2$	0.487	−2.084
4c	O	H	H	H	H	H	$C_{27}H_{19}NO_2$	0.553	−1.854
4d	S	CH_3	H	i-Pr	H	CH_3	$C_{32}H_{29}NS_2$	0.338	−2.065
4e	S	H	CH_3	H	CH_3	CH_3	$C_{30}H_{25}NS_2$	0.470	−2.090
4f	S	H	H	H	H	H	$C_{27}H_{19}NS_2$	0.567	−1.858

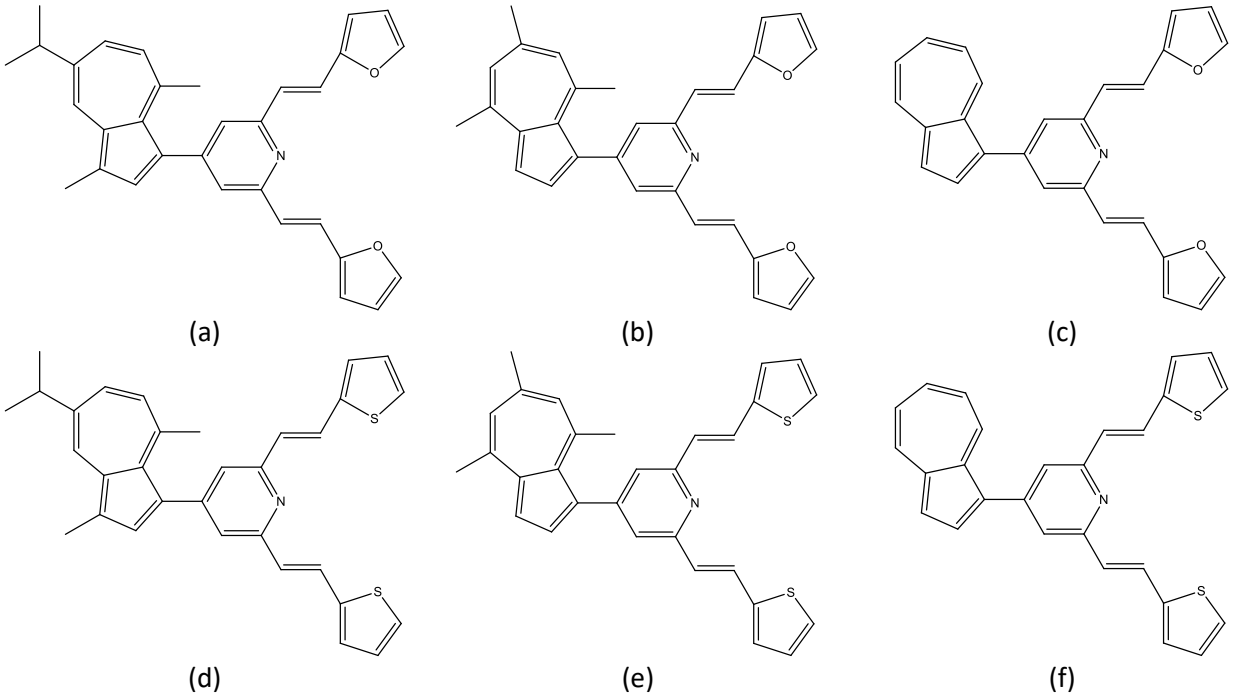


Figure 4. *4-(azulen-1-yl)-2,6-divinylpyridine derivatives*

Molecular descriptors families

Physical, chemical and biological activities of chemical compounds are related to one another in a natural way since all molecules are constituted from atoms and the same forces keep the atoms together in molecules. It is thus a supported idea to construct a pool of molecular descriptors that are closely related and differ by only a choice in the construction. A series of such pools of descriptors was reported (FPIF in Chapter 7 of [28]; MDF in [29]; MDFV in [30]; SAPF in [31]; SMPI in [32], FMPI in [33]), and some notable results in the modeling of biological activities were obtained [34-37]. A thesis [38] elaborates on the issue.

For convenience, Table 3 provides the coding of the FMPI.

Table 3. *Templating of FMPI*

Gene	L_O					M_O					I_D				D_M			A_P							F_C		
Genome	I	R	L	E	F	I	J	N	M	D	E	P	U	G	T	U	A	B	C	D	E	F	G	M	N	S	

Code: FMPI = LOMOIDDM AP FC; Descriptors: 4536

The encodings from Table 3 correspond to:

- Fragmentation criteria (of the molecule) F_C with maximal fragments for F_C = M, with minimal fragments for F_C = N, and with Szeged fragments for F_C = S;
- Atomic property A_P with (fragment's) sum of atomic masses (in Da) for A_P = A, harmonic sum of atomic numbers (Z) for A_P = B, with sum of cardinalities for A_P = C, with harmonic mean of solid state density (in kg/m^3) for A_P = D (see page 12 in [39]), with geometrical mean of electro-negativities from revised Pauling scale for A_P = E (see page 16 in [39]), with average of the first ionization energy (in kJ/mol) for A_P = F (see page 14 in [39]) and with power 2 mean (PM(p), p = 2) of the melting point temperatures (K) for A_P = G (see page 10 in [39]);
- Distance metric D_M with (in between atoms) geometric (in nm) for D_M = G, topologic (in bonds) for D_M = T, and weighted (by bond order) topologic for D_M = U; interaction descriptor I_D (built with selected property and distance) with $P_{i,j}D_{i,j}$ for I_D = E, with $P_{i,j}/D_{i,j}$ for I_D = U, with $P_{i,j}$ for I_D = P and with $D_{i,j}$ for I_D = D;

- Mean operation (averaging from multiple fragments of a molecule) M_O, with min.($M_{i,j}$) for M_O = N, max. ($M_{i,j}$) for M_O = M, $^1/_2\Sigma(M_{i,j})$ for M_O = I, $^1/_2\Sigma(M_{i,j}M_{j,i})$ for M_O = J, $^1/_2\Sigma(M_{i,j}Ad_{i,j})$ for M_O = E, $^1/_2\Sigma(M_{i,j}M_{j,i}Ad_{i,j})$ for M_O = F;
- Linearization operator (adjusting the scale of the descriptor) L_O with $f(x) = x$ for L_O = I, with $f(x) = x^{-1}$ for L_O = R, and with $f(x) = \ln x$ for L_O = L.

Extending the characteristic polynomial

The characteristic polynomial (ChP) was first expressed in the 18th century in order to characterize the movement of planets [40]. Later, Hückel's method of molecular orbitals [41] was the ChP first extension approximating the treatment of π electron systems in organic molecules. An extended ChP (EChP) in a parameter (x in Equation (1)) can be associated with two particular matrices (square, of size n; A and I in Equation (1)) through the evaluation of a determinant [42]:

$$\text{EChP}(x,I,A) \leftarrow |xI-A| \quad (1)$$

In Equation (1), I should be a diagonal matrix (a matrix in which the entries outside the main diagonal are all zero), while A should be the exact opposite (a matrix from which the entries of the main diagonal were extracted). In this instance, I and A can be associated with an (edge) directed and (vertex) labeled graph with no loops. The undirected and unlabeled graph is the case that recovers the classical formula of the ChP (for instance, the one applied in [43]).

One should notice that the classical formula of the ChP is not well suited for modeling the considered congeners (Figure 4) since different chemical elements are present (nitrogen, oxygen and sulfur, in addition to carbon), while an EChP formula will take into consideration the atoms' change.

For convenience, Table 4 provides the coding of the EChP. The encodings from Table 4 correspond to (see also [42] and [44]):

Table 4. *Templating of EChP*

Gene	L_O			I_A								A_C				d_0				d_1, d_2, d_3									
Genome	I	R	L	A	B	C	D	E	F	G	H	t	g	c	T	G	C	-	+	0	1	2	3	4	5	6	7	8	9

Code: EChP = $L_OI_AA_C(d_0.d_1d_2d_3)$; Descriptors: 288144

- Atomic identity, I_A (variable I in Equation (1)), with A = A for atomic mass divided to 294.0 (atomic mass of Oganesson), A = B for cardinality (always 1), A = C for electrostatic charges (ESP method [45]), A = D for solid-state density (in kg/m^3, divided to 30000; see page 12 in [39]), A = E for electronegativity (revised Pauling, divided to 4.00, see page 16 in [39]), A = F for first ionization potential (in kJ/mol, /1312.0, see page 14 in [39]), A = G for melting point temperature (in K, divided to 3820.0, see page 10 in [39]), and A = H for the number of attached hydrogen atoms (divided to 4, the valence of a carbon atom);
- Connectivity parameter A_C (variable A in Equation (1)), with connectivity on adjacencies (classical ChP, [A] [Ad], topological adjacency matrix) for C = t, connectivity on topological distances (classical ChP, [A] [Di], topological distance matrix) for C = T, inverse of the geometrical distance for C = G, and only for adjacent atoms for C = g, inverse of the conventional bond orders for C = c, and inverse of the sum of conventional bond orders for C = C;
- Evaluation point (variable x in Equation (1)), expressed with four digits, as $\pm\overline{0.d_1d_2d_3}$, with d_1, d_2, $d_3 \in \{0, 1, ..., 9\}$;
- Linearization operator (adjusting the scale of the descriptor) L_O with $f(x) = x$ for O = I, with $f(x) = x^{-1}$ for O = R, and with $f(x) = \ln x$ for O = L.

Quantitative structure-property relationship studies

For a biologist, basic units of structure define the function of all living things [46]. Since 1868, when Crum-Brown and Fraser argued about the existence of the relation between the physiological action of a substance and its composition and constitution [47], much progress has been made. Today, structure-property relationships may be considered the central dogma of chemistry [48].

A typical structure-property study (referred to as quantitative, quantitative structure-property relationship (QSPR) to express its desired outcome, a relationship expressing the property as a function of the structure) requires the collection of a series of compounds with known structure and measured properties.

Certain inclusion criteria apply. Therefore, the presence of an extreme value [49] or an outlier [50] in the series makes the derived relations unreliable, so the derivation of the structure-property relationships is desired to be obtained in the absence of those. The environment, such as water as a solute, may influence dissociations and associations of ions [51], and chemical properties [26] are also affected. Furthermore, it should be a clearly defined endpoint, an unambiguous algorithm to facilitate the reproducibility of a built model, a defined applicability domain to determine the model space associated with reliable predictions, to employ appropriate measures of goodness-of-fit, robustness and predictivity, and ideally, a mechanistic interpretation of the model in chemical structural terms [52].

One may argue that modeling of the 3D geometry is of the essence for estimating the biological activities [53], while, for physical properties, feeding the models with topologies works well (take, for instance log *P* [54]).

Two methods were employed: with FMPI (Table 3) and with EChP (Table 4), considering the molecules' topology and geometry. As expected, the 3D model of the computed molecular geometry depends on the selected computational model. Three models were subjected to analysis: a mechanistic one (MMFF), an in vitro one (B3LYP), and a solvent-interacting one (M06 SM8). The design of this experiment is depicted in Table 5.

Table 5. *Molecular methods for optimal conformations*

Design	Molecular modeling method	Ref.
Raw	MMFF94	[55]
In vitro	B3LYP 6-311G**	[56]
In vivo	M06 6-31G* + SM8-water	[57]

See Table 6 in Supplementary material of [56] for the reasons of choosing B3LYP

Using notation from [58], if $f = f(\underline{x}; b)$ depending on x as the random variable and having $b = (b_j)_{1\le j\le m}$ as m population parameters) is the theoretical distribution function, then the likelihood of the sample is $f(x_1;b)f(x_2;b)\ldots f(x_n;b)$, and it is maximum when all its derivatives vanish. The maximum likelihood estimation (MLE) method was used here.

Results and discussion

Modeling the oxidation and reduction potential of vynil-pyridines

There are six observation points for each electrochemical property in Table 3 (n = 6), which is, at the minimum required to conduct an association analysis. However, in [59], it is pointed out that at n < 8, both false positives and false negatives may appear if the association is used to do a classification (which is not the case here). At the same time, other studies (including Google's AI) suggest at least n = 10. A sample size calculation can be employed [60], but it is useful only when sampling is available. Therefore, this deficiency must be alleviated by better testing the sample's quality.

Outlier and extreme values analysis was conducted. Normal distribution ($\mathcal{N}$) population parameters (μ and σ), associated with the sample for the given measurements of the properties, were obtained using MLE and are given in Table 6. One should note that the usage of MLE, unlike the method of central moments, does not reduce the number of degrees of freedom associated with the sample [61].

Table 6. *Gaussian distribution population parameters of the 4-(azulen-1-yl)- 2,6-divinylpyridine derivatives by maximizing the likelihood for the samples from Table 2*

Potential \Parameters	μ	σ
Oxidation (E_a)	0.4555	0.10569
Reduction (E_c)	-2.0037	0.11474

μ: mean; σ: standard deviation; σ^2: variance

The results in Table 6 have been used to obtain the cumulative probabilities associated with each observation. Table 7 contains these results.

Table 7. *Cumulative probabilities (data in Table 2; parameters in Table 6)*

i	1	2	3	4	5	6
Oxidation potentials (sorted data)						
x_i	0.318	0.3380	0.47	0.487	0.5530	0.567
p_i	0.0966	0.1331	0.5546	0.6172	0.8219	0.8543
Reduction potentials (sorted data)						
x_i	-2.09	-2.084	-2.071	-2.065	-1.858	-1.854
p_i	0.226	0.242	0.2788	0.2966	0.8979	0.904

$p_i = \int_{x \le x_i} f(x;b)dx$ are the cumulative probabilities

A series of order statistics was calculated. Table 8 contains these results. The analysis from Table 8 reveals that there are 99.4 % worse draws of a normal ($\mathcal{N}$) sample than the sample of E_a values and 44.5 % worse draws of a normal ($\mathcal{N}$) sample than the sample of E_c values. Both probabilities (p values in boldface in Table 8) are greater than the conventional level of 5 %, so the assumption that the samples of E_a values and E_c values are drawn from normal distributions ($\mathcal{N}$) cannot be rejected.

Table 8. *Order statistics on rejecting the hypothesis of samples normality*

Statistic	CM	WU	KS	KV	AD	H1	g1	TS	FCS
k	1	2	3	4	5	6	7	8	
Statistics and associated probabilities for the series of observed E_a values									
Value	0.36302	0.54199	0.05696	1.03250	0.05596	2.94610	0.40337	0.29957	2.44592
p_k	0.829	0.875	0.857	0.606	0.710	0.564	0.724	0.792	0.994
Statistics and associated probabilities for the series of observed E_c values									
Value	0.87742	0.90647	0.15756	1.47290	0.15358	2.93340	0.40400	0.29486	8.88016
p_k	0.422	0.304	0.373	0.096	0.089	0.575	0.722	0.818	0.445

FCS: Fisher's chi-square statistic; $ECS = -\ln\sum_{k=1}^{8} p_k$; $p_{FCS} = 1 - CDF_{\chi^2}(FCS, 8)$

Modeling with raw geometries

Table 9 contains step-by-step results for modeling with SAPF and EChP using optimum geometries from the MMFF94 molecular mechanics approach. One should note that the descriptors from EChP are able to better explain both E_a and E_c electrochemical properties than the ones from FMPI (see r^2_{adj} values in Table 9).

Table 9. *Staged selection of the best descriptors for linear associations with the descriptors from raw geometries*

Family	Stage	For E_a dataset	For E_c dataset	Result
FMPI	1	2275	2275	Number of adapted descriptors
FMPI	2	2069	2198	Number of linearly associated descriptors
FMPI	2	LNUTGM	INEGEM	Best performing descriptors
FMPI	2	0.9894	0.9949	r^2 with above descriptors
EChP	1	265411	265411	Number of adapted descriptors
EchP	2	252120	243742	Number of linearly associated descriptors
EChP	2	RDCN0940	LEGN0705	Best performing descriptors
EChP	2	0.9977	0.9996	r^2 with above descriptors

RDCN0940 = (EChP(−0.940; I_D, A_C))$^{-1}$, LEGN0705 = ln(EChP(−0.705; I_E, A_G)) in Equation (1) and Table 4

With MMFF94 obtained geometries, for the considered 4-(azulen-1-yl)-2,6-divinylpyridine derivatives, from the EChP selected descriptors, it can be inferred that the solid-state density is selected as the influential property for E_a acting on the inverse of the sums of conventional bond orders ($I \leftarrow I_D$ and $A \leftarrow A_C$ extending the ChP, see Equation 4.1 and Table 9). On the other hand, E_c is expressed by the electronegativity acting on the inverse of the geometrical distance ($I \leftarrow I_E$ and $A \leftarrow A_G$ extending the ChP, see Equation 4.1 and Table 9). However, FMPI, with only 1.6 % of the size of EChP, explains over 99 % of the variance in the E_a and E_c values and can be used as a first approximation level, just as MMFF is used for molecular geometry. Going further with the analysis of the FMPI selected descriptors, E_a is related to the structure on a logarithmic scale (”L” letter in LNUTGM), while E_c is related on an identity scale (”I” letter in INEGEM), both E_a and E_c were seen through FMPI perspective as localized properties (properties characterized by minimal fragments, usually single atoms - ”N” as the second letter in LNUTGM and INEGEM descriptors’ names, see Table 9 and Table 3 encodings), interacting like a potential (property over distance ”U” as third letter in the LNUTGM name, see Table 9 and Table 3 encodings) and like a elastic force (property multiplied with the distance ”E” as third letter in the INEGEM name, see Table 9 and Table 3 encodings) respectively, being affected mainly by the topology of the molecules (”T” as third letter in LNUTGM), and by the geometry (”G” as third letter in INEGEM) respectively. The differences are not at the level of the fragmentation criteria (for both E_a and E_c is minimal, ”M” as last letter in LNUTGM and INEGEM), but at the elemental property relating the best (for E_a is the melting point temperature, ”G” letter in LNUTGM, and for E_c is the electronegativity, ”E” before last letter in INEGEM).

Table 10 lists the calculated values of the descriptors for each molecule from the set.

Table 10. *Calculated values of the descriptors from raw geometries*

Descriptor	Family		Compounds from Figure					
			4a	4b	4c	4d	4e	4f
LNUTGM	FMPI	Values estimating E_a	7.326	7.367	7.392	7.326	7.367	7.393
INEGEM	FMPI	Values estimating E_c	3.426	3.425	3.432	3.426	3.425	3.432
-10^5RDCN0940	EChP	Values estimating E_a	15.21	6.158	2.642	14.37	7.003	2.477
-10^1LEGN0705	EChP	Values estimating E_c	3.608	3.677	2.775	3.581	3.685	2.781

10^5 and 10^1 are scaling factors; Conventionally slope and intercept from Table 11 have units of V; then here values are adimensional

One can notice that INEGEM descriptor from FMPI does not distinguish between Oxygen and Sulphur (furan and thiophen rings in Figure 4), LNUTGM barely distinguishes between 4c and 4f, while the EChP descriptors make the distinction (see the numeric values in the 4a, 4b, 4c group *vs.* 4d, 4e, 4f group in Table 10). The obtained linear models expressing E_a and E_c as functions of derivatives structure are provided in Table 11.

Table 11. *Models and associated statistics with the descriptors from raw geometries*

Property	Family	Model	p_F
E_a	FMPI	$y = -25.4_{\pm 3.3} + 3.51_{\pm 0.45}$ LNUTGM	3×10^{-5}
E_a	EChP	$y = 0.606_{\pm 0.011} + 1.89_{\pm 0.11} \cdot 10^{-7} \cdot (\mathrm{EChP}(-0.940_{\pm 0.0005}; I_A = D, A_C = C))^{-1}$	10^{-6}
E_c	FMPI	$y = -118_{\pm 10} + 33.8_{\pm 3.0}$ INEGEM	6×10^{-6}
E_c	EChP	$y = -1.141_{\pm 0.022} + 0.026_{\pm 0.01} \cdot \ln(\mathrm{EChP}(-0.705_{\pm 0.0005}; I_A = E, A_C = G))$	4×10^{-8}

$y \sim \mathrm{lin}(x)$, $y \in \{E_a, E_c\}$, $x \in$ FMPI or $x \in$ EChP, $\mathrm{lin}(x) \in \{c_0x, c_1x + c_2\}$, $c_0, c_1, c_2 \in \Re^*$; slope and intercept have units of V

In both instances (for E_a and E_c), with the use of the EChP approach, the precision of the models is improved.

Modeling with in vitro geometries.

Table 12 contains step-by-step results of modeling with SAPF and EChP using optimum geometries from B3LYP 6-311G** theory level.

Table 12. *Staged selection of the best descriptors for linear associations with descriptors from in vitro geometries*

Family	Stage	For E_a dataset	For E_c dataset	Result
FMPI	1	2218	2218	Number of adapted descriptors
FMPI	2	2049	2041	Number of linearly associated descriptors
FMPI	2	LNUTGM	LNEGCM	Best performing descriptors
FMPI	2	0.9894	0.9909	r^2_{adj} with above descriptors
EChP	1	265750	265750	Number of adapted descriptors
EChP	2	247097	238368	Number of linearly associated descriptors
EChP	2	RDCN0940	ICtN0193	Best performing descriptors
EChP	2	0.9977	0.9995	r^2_{adj} with above descriptors

RDCN0940 = $(\mathrm{EChP}(-0.940; I_A = D, A_C = C))^{-1}$, ICtN0193 = $\mathrm{EChP}(-0.193; I_A = C, A_C = t)$ in Equation (1)

One should note that the descriptors from EChP are able to better explain both E_a and E_c electrochemical properties than the ones from FMPI (see r^2 values in Table 12).

With B3LYP obtained geometries, for the considered 4-(azulen-1-yl)-2,6-divinylpyridine derivatives, from the EChP selected descriptors, can be inferred that the solid-state density is selected as the influential property for E_a acting on the inverse of the sums of conventional bond orders ($I \leftarrow I_D$ and $A \leftarrow A_C$ extending the ChP, see Equation (1) and Table 9). On the other hand, E_c is expressed by the electrostatic charges acting on adjacencies ($I \leftarrow I_C$ and $A \leftarrow$ *)* extending the ChP, see Equation (1) and Table 12).

Through the perspective of FMPI selected descriptors, E_a and E_c are related with the structure on a logarithmic scale (”L” letter in LNUTGM and LNEGCM), both E_a and E_c were seen through FMPI perspective as localized properties (properties characterized by minimal fragments, usually single atoms - ”N” as second letter in LNUTGM and LNEGCM descriptors names, see Table 9 and Table 3 encodings), interacting like a potential (property over distance - ”U” as third letter in LNUTGM name, see Table 9 and Table 3 encodings) and like an elastic force (property multiplied with the distance - ”E” as third letter in LNEGCM name, see Table 9 and Table 3 encodings) respectively, being affected mainly by the topology of the molecules (”T” as third letter in LNUTGM), and by the geometry (”G” as third letter in LNEGCM) respectively. The differences are not at the level of the fragmentation criteria (for both E_a and E_c, it is minimal, ”M” as last letter in LNUTGM and INEGEM) but at the elemental property relating the best (for E_a is the melting point temperature, ”G” letter in LNUTGM, and for E_c is the cardinality (size of the fragment), ”C” before last letter in LNEGCM).

Table 13 lists the calculated values of the descriptors for each molecule from the set.

Table 13. *Calculated values of the descriptors from in vitro geometries*

Descriptor	Family		Compound					
			4a	4b	4c	4d	4e	4f
LNUTGM	FMPI	Values estimating E_a	7.326	7.367	7.392	7.326	7.367	7.393
LNEGCM	FMPI	Values estimating E_c	0.4065	0.4170	1.702	0.4064	0.4166	1.708
-10^5RDCN0940	EChP	Values estimating E_a	15.21	6.158	2.642	14.37	7.003	2.477
-10^0ICtN0193	EChP	Values estimating E_c	0.1515	1.054	−10.3	.1530	1.142	−10.1

10^5 and 10^0 are scaling factors; Conventionally slope and intercept from Table 14 have units of V; then here values are adimensional

One can notice that the INEGEM descriptor from FMPI does not distinguish between oxygen and sulphur (furan and thiophene rings in Figure 4), LNUTGM barely distinguishes between 4c and 4f, while the EChP descriptors make the distinction (see the numeric values in the 4a, 4b, 4c group *vs.* 4d, 4e, 4f group in Table 13).

The obtained linear models expressing E_a and E_c depending on the structure of derivatives are provided in Table 14.

Table 14. *Models and associated statistics with the descriptors from in vitro geometries*

Property	Family	Model	p_F
E_a	FMPI	$y = -25.4_{\pm 3.3} + 3.51_{\pm 0.45} \cdot$LNUTGM	3×10^{-5}
E_a	EChP	$y = 0.606_{\pm 0.011} + 1.89_{\pm 0.11} \cdot 10^{-7} \cdot (\text{EChP}(-0.940_{\pm 0.0005}; I_D, A_C))^{-1}$	10^{-6}
E_c	FMPI	$y = -2.15_{\pm 0.02} + 0.17_{\pm 0.02} \cdot$LNEGCM	2×10^{-5}
E_c	EChP	$y = -2.065_{\pm 0.003} - 0.0205_{\pm 0.0006} \cdot \text{EChP}(-0.193_{\pm 0.0005}; I_C, A_t)$	6×10^{-8}

$y \sim \text{lin}(x)$, $y \in \{E_a, E_c\}$, $x \in$ FMPI or $x \in$ EChP, $\text{lin}(x) \in \{c_0x, c_1x + c_2\}$, $c_0, c_1, c_2 \in \Re^*$; slope and intercept have units of V

In both instances (for E_a and E_c), with the use of the EChP approach, the precision of the models is improved.

Modeling with in vivo *(water) geometries.*

Table 15 contains step-by-step results of modeling with SAPF and EChP using optimum geometries from M06 6-31G* + SM8-water theory level.

Table 15. *Staged selection of the best descriptors for linear associations*

Family	Stage	For E_a dataset	For E_c dataset	Result
FMPI	1	2220	2220	Number of adapted descriptors
FMPI	2	2050	2052	Number of linearly associated descriptors
FMPI	2	LNUTGM	RNUTCM	Best performing descriptors
FMPI	2	0.9894	0.9923	r^2_{adj} with above descriptors
EChP	1	265964	265964	Number of adapted descriptors
EChP	2	251531	243044	Number of linearly associated descriptors
EChP	2	RDCN0940	ICtN0176	Best performing descriptors
EChP	2	0.9977	0.9994	r^2_{adj} with above descriptors

RDCN0940 = $(\text{EChP}(-0.940; I_D, A_C))^{-1}$, ICtN0176 = $\text{EChP}(-0.176; I_C, A_t)$ in Eq. 1

One should note that the descriptors from EChP are able to explain better than the ones from FMPI, both E_a and E_c electrochemical properties (see r^2 values in Table 15).

With M06 6-31G* + SM8-water obtained geometries, for the considered 4-(azulen-1-yl)-2,6--divinyl-pyridine derivatives, the solid-state density is selected as the influential property for E_a, acting on the inverse of the sums of conventional bond orders ($I \leftarrow I_D$ and $A \leftarrow A_C$ extending the ChP, see Equation 4.1 and Table 9). On the other hand, E_c is expressed by the electrostatic charges acting on adjacencies ($I \leftarrow I_C$ and $A \leftarrow A_t$ extending the ChP, see Equation and Table 15). However, FMPI,

with only 1.6 % of the size of EChP, explains over 99 % of the variance in the E_a and E_c values and can be used as a first approximation level, just as MMFF is used for molecular geometry. Going further with the analysis of the FMPI selected descriptors, E_a is related to the structure on a logarithmic scale (”L” letter in LNUTGM) while E_c is related on a reciprocal scale (”R” letter in RNUTCM), both E_a and E_c were seen through FMPI perspective as localized properties (properties characterized by minimal fragments, usually single atoms - ”N” as second letter in LNUTGM and RNUTCM descriptors names, see Table 9 and Table 3 encodings), interacting like a potential (property over distance - ”U” as third letter), being affected mainly by the topology of the molecules (”T” as third letter). The differences are not at the level of the fragmentation criteria (for both E_a and E_c it is minimal, ”M” as last letter in LNUTGM and RNUTCM), but at the elemental property relating the best (for E_a is the melting point temperature, ”G” letter in LNUTGM, and for E_c it is the cardinality (size of the fragment), ”C” before last letter in RNUTCM).

Table 16 lists the calculated values of the descriptors for each molecule from the set.

Table 16. *Calculated values of the descriptors from in vivo geometries*

Descriptor	Family		Compdound					
			4a	4b	4c	4d	4e	4f
LNUTGM	FMPI	Values estimating E_a	7.326	7.367	7.392	7.326	7.367	7.393
RNUTCM	FMPI	Values estimating E_c	1.000	1.000	0.7143	1.000	1.000	0.7143
-10^5RDCN0940	EChP	Values estimating E_a	15.21	6.158	2.642	14.37	7.003	2.477
-10^1ICtN0176	EChP	Values estimating E_c	−0.7679	8.225	−104.0	−0.8165	8.454	−102.4

10^5 and 10^1 are scaling factors; Conventionally slope and intercept from Table 17 have units of V; then here values are adimensional

One can notice that the best descriptors from FMPI do not distinguish between 4a and 4c and between 4b and 4d, while EChP does.

The obtained linear models expressing E_a and E_c as functions of derivatives structure are provided in Table 17.

Table 17. *Models and associated statistics*

Property	Family	Model	p_F
E_a	FMPI	$y = -25.4_{\pm 3.3} + 3.51_{\pm 0.45} \cdot$LNUTGM	$3 \cdot 10^{-5}$
E_a	EChP	$y = 0.606_{\pm 0.011} + 1.89_{\pm 0.11} \cdot 10^{-7} \cdot (\text{EChP}(-0.940_{\pm 0.0005}; I_D, A_C))^{-1}$	$1 \cdot 10^{-6}$
E_c	FMPI	$y = -1.302_{\pm 0.077} - 0.775_{\pm 0.085} \cdot$RNUTCM	$1 \cdot 10^{-5}$
E_c	EChP	$y = -2.0697_{\pm 0.0038} + 0.0207_{\pm 0.0006} \cdot \text{EChP}(-0.176_{\pm 0.0005}; I_C, A_t)$	$9 \cdot 10^{-8}$

$y \sim \text{lin}(x)$, $y \in \{E_a, E_c\}$, $x \in$ FMPI or $x \in$ EChP, $\text{lin}(x) \in \{c_0x,\ c_1x + c_2\}$, $c_0, c_1, c_2 \in \Re^*$; slope and intercept have units of V

In both instances (for E_a and E_c), with the use of the EChP approach, the precision of the models is improved.

Correlated correlations analysis

Correlated correlation analysis is intended to reveal the change in the information carried by a model. The main supposition is that when changing a model, the same data was used to derive the new model, such that the association in the paired data still exists. In a classical analysis, the correlation coefficient between the experimental and the calculated is of interest. With two models: Model 1, with $y \sim \text{lin}(x_1)$, $r_{01} = r(y, x_1)$, and Model 2, with $y \sim \text{lin}(x_2)$, $r_{02} = r(y, x_2)$, the calculation of a third correlation coefficient, $r_{12} = r(x_1, x_2)$, may provide the necessary information to reveal the change in the information carried by $y \sim \text{lin}(x_1)$ model *vs.* $y \sim \text{lin}(x_2)$ model. Steiger [62] proposed

the use of such a test. Its approach is used here. Equation (2) has been used (conventionally positive correlations have been used, $r \leftarrow |r|$):

$$S = \left(Z_{01} - \frac{r_{01}}{2(n-1-f_1)} - Z_{01} + \frac{r_{01}}{2(n-1-f_1)} \right) \sqrt{\frac{(n-3)(1-r_a^2)^2}{2-3r_a^2-\rho_{12}(2-4r_a^2+r_a^2\rho_{12})}} \quad (2)$$

where Z_{01}, Z_{02}, and Z_{12} are the Fisher's Z transformations ($Z(r) = 0.5(\ln(1+r)/\ln(1-r))$) of the correlation coefficients (r_{01}, r_{02}, and r_{12}), $r_a = (r_{01} + r_{02})/2$, $\rho_{12} = (1\text{-}e^{2o_{12}})/(1+e^{2o_{12}})$, $o_{12} = Z_{12} - 0.5r_{12}/(n-1-f_m)$, and where f_1, f_2, and $f_m = (f_1 + f_2)/2$ are the degrees of freedom associated with the models. The probability is calculated from Student's t distribution with $n-2$ degrees of freedom ($p_S = \text{CDF}_t(|S|; n-2)$).

Correlated correlation analysis is a powerful test, and when the model is changed, probabilities to gain a different model can be calculated. Thus, Table 18 contains the probabilities associated with the changes in the models.

Table 18. *Correlated correlation analysis distinguishing between the models*

Fixed	Changing	y	x_1	x_2	S	$p_t(S; 2)$
raw	FMPI *vs* EChP	E_a	LNUTGM	RDCN0940	1.32	0.3177
raw	FMPI *vs* EChP	E_c	RNUTAN	LEGN0705	5.56	0.0309
vitro	FMPI *vs* EChP	E_a	LNUTGM	RDCN0940	1.32	0.3177
vitro	FMPI *vs* EChP	E_c	LNEGCM	ICtN0193	2.52	0.1279
vivo	FMPI *vs* EChP	E_a	LNUTGM	RDCN0940	1.32	0.3177
vivo	FMPI *vs* EChP	E_c	RNUTCM	ICtN0176	2.18	0.1611
EChP	raw *vs* vitro	E_c	LEGN0705	ICtN0193	0.19	0.8668
EChP	raw *vs* vivo	E_c	LEGN0705	ICtN0176	0.35	0.7598
EChP	vitro *vs* vivo	E_c	ICtN0193	ICtN0176	0.75	0.5315

Geometry - optimum from: raw - MMFF94; vitro - B3LYP 6-311G**; vivo - M06+Sm8-water 6-31G*;
structure descriptors - families: fragments based - FMPI; ChP extended - EChP;
electrochemical - potentials: E_a - oxidation; E_c - reduction; S: Steiger's statistic calculated with Equaation (2) ($r_{01} = r(y, x_1)$, $r_{02} = r(y, x_2)$, $r_{12} = r(x_1, x_2)$); p_t: probability that the samples of the changing parameters to be drawn from the same population; $p_t(S; 2)$ for RNUTAN *vs* LEGN0705 is statistically significant (< 5 %; in red color)

There are several models provided here (all statistically significant). Table 11 contains four models, two each for the prediction of E_a and E_c with each family of descriptors. The first two models from Table 11 are not statistically distinct one from the other (the probability of expressing the same structural information is 31.77 %, see Table 18). The last two models from Table 11 are statistically distinct from each other (the probability of expressing the same structural information is 3.09 %, see Table 18) and are the only ones in distinction, according to the results from Table 8. Table 18 supports the idea that LEGN0705 (expressing E_c from files containing raw geometries), ICtN0193 (expressing E_c from files containing in vitro geometries), and ICtN0176 (expressing E_c from files containing *in vivo* geometries) may essentially express the same information relating structure - reduction potential since statistical significance has not been reached to distinguish between them. Furthermore, results listed in Tables 11, 14, and 17 reveal that, when from files containing raw, in vitro and in vivo geometries, E_c is expressed with EChP descriptors, there is no change in the selection of the model best expressing the association (that is the model containing RDCN0940 molecular descriptor).

There is no coincidence in the fact that an extended family of descriptors based on the characteristic polynomial (EChP) performs better even than a fragmentation-based method (FMPI) in relating the chemical structure with electrochemical properties (here, electrode potentials of 4-(azulen-1-yl)-2,6-divinylpyridine derivatives), since the roots of the eigenproblem run deep in chemistry [63].

Conclusions

The present study has shown a significant gain in the precision of expressing the QSPR model coefficients when EChP descriptors are used instead of FMPI descriptors. At the same time, EChP descriptors have a better resolution - they distinguish between all molecules from the dataset, while the resolution of the best explanatory FMPI descriptors did not distinguish between furan and thiophene rings, and this is a positive result.

The use of higher theory-level models in geometry optimization (from MMFF94 to B3LYP and M06 + Water SM8) did not obtain the expected gain in precision and/or resolution, which is a negative result. In the case of EChP model of E_a, the selected descriptor remained the same (RDCN0940) and its values remained unchanged, while in the case of the EChP model of E_c, the selected descriptor has been changed (LEGN0705, to ICtN0193 and to ICtN0176), but not with a supplementary precision or resolution. On the contrary, a slow loss of precision was the result (r^2 from 0.9996 to 0.9995, and to 0.9994; p_F from 4×10^{-8} to 6×10^{-8}, and to 9×10^{-8}).

MMFF94 geometries proved better suited to express the E_a and E_c potentials from structure than the selected B3LYP and M06 upper theory level approaches. This is a positive result regarding MMFF94 (one should notice that this is the default theory level in providing geometries of molecules by PubChem), and perhaps supports the idea that selecting and using upper theory levels in molecular geometry optimization requires much more care and suitability analysis than expected.

Correlated correlation analysis revealed an important fact. The EChP best model (containing the LEGN0705 descriptor) is statistically distinct (evidence in Table 18), and better (evidence in Table 9) than the FMPI best model (containing the INEGEM descriptor) in expressing structure-property relationship for E_c.

Abbreviations

CAS RN: Chemical Abstracts service registry number
PubChem: http://pubchem.ncbi.nlm.nih.gov/ - chemical compounds search engine and database online available
CID: PubChem compound identifier
FPIF: fragmental property index family
MDF: molecular descriptors family
SAPF: structural atomic property family
SMPI: Szeged matrix property indices
FMPI: fragmental matrix property indices
i-Pr: isopropyl
ChP: characteristic polynomial
EChP: extended characteristic polynomial
QSPR: quantitative structure-property relationships
DFT: density functional theory
MLE: Maximum Likelihood Estimation
CDF: cumulative distribution function
PDF: probability density function
r: Pearson's correlation coefficient [64]
r^2_{adj}: adjusted (from *r*) determination coefficient
χ^2: chi square distribution; $PDF_{\chi^2}(x; k) = x^{k/2-1}e^{-x/2}2^{-k/2}(\Gamma(k-2))^{-1}$
$\mathcal{N}$: normal distribution; $PDF_{\mathcal{N}}(x; \mu, \sigma) = \sigma^{-1}(2\pi)^{-1/2}e^{-(x-\mu)^2/(2\sigma^2)}$

t: Student's t distribution [65]; $PDF_t(x; \nu) = (\pi\nu)^{-1/2}\Gamma\left(\frac{\nu+1}{2}\right)\left(\Gamma\left(\frac{\nu}{2}\right)\right)^{-1}\left(1+\frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$

$\Re^*$: the set of real non-null numbers ($R^* = R\backslash 0$)

MMFF: Merk molecular force field; MMFF94: 1994 version of MMFF [55]

B3LYP: Becke 3 parameter hybrid exchange functional and Lee-Yang-Par correlation functional [66, 67]

M06: Minnesota density functional introduced in [68]

SM8: solvation model applicable to all solvents introduced in [69]

p_F : probability of a wrong model, calculated from Fisher's F distribution [70]

Appendix: Generated data and data analysis

Compounds studied in the paper are not available in the PubChem database. We drawn their 2D structures and build up their 3D models. Chemdraw files are available at: http://lori.academicdirect.ro/data/C19H15N-d/Data_Chemdraw.zip.

The 3D geometries were build (with HyperChem v.8 Add H & Model Build menu option and module), exported and optimized into Spartan (v. 14) software (using three theory levels, as described in the paper). Further, structure-activity relationship search was conducted using two approaches (FMPI and EChP), as described in the paper.

The subsequent analysis is available at:
http://lori.academicdirect.ro/data/C19H15N-d/Results_MMFF94.zip (the analysis with MMFF94 optimum geometries, big data file, over 30 MB),
http://lori.academicdirect.ro/data/C19H15N-d/Results_MMFF94.zip (the analysis with B3LYP optimum geometries, big data file, over 40 MB), and
http://lori.academicdirect.ro/data/C19H15N-d/Results_MMFF94.zip (the analysis with M06 optimum geometries, big data file, about 40 MB).

For the reproducibility of the results, the programs are provided as well. Thus, the programs supporting the study with EChP are available at:
http://lori.academicdirect.ro/data/C19H15N-d/Programs_EChP.zip,
the programs supporting the study with FMPI are available at:
http://lori.academicdirect.ro/data/C19H15N-d/Programs_FMPI.zip
and the programs filtering statistically the structure-based descriptors are available at:
http://lori.academicdirect.ro/data/C19H15N-d/Programs_LR.zip.

One need to prepare (inspect and modify by necessity) the properties.asc file listing the experimenttal properties in order to use the methodology on other dataset. Then the programs should be called in order.

References

[1] M. D. Hill. Recent strategies for the synthesis of pyridine derivatives. *Chemistry - A European Journal* **16** (2010) 12052-12062. http://doi.org/10.1002/chem.201001100

[2] C. Pang, Y. Xu, X. Ma, S. Li, S. Zhou, H. Tian, M. Wang, B. Han. Design, synthesis, and evaluation of novel arecoline-linked amino acid derivatives for insecticidal and antifungal activities, *Scientific Reports* **14** (2024). http://doi.org/10.1038/s41598-024-60053-2

[3] M. Feng, B. Gao, D. Ruiz, L. R. Garcia, Q. Sun. Bacterial vitamin b6 is required for postembryonic development in c. elegans, *Communications Biology* **7** (2024). http://doi.org/10.1038/s42003-024-05992-2

[4] M. Kang, H. Wang, C. Chen, R. Suo, J. Sun, Q. Yue, Y. Liu. Analytical strategies based on untargeted and targeted metabolomics for the accurate authentication of organic milk from jersey and yak, *Food Chemistry X* **19** (2023). http://doi.org/10.1016/j.fochx.2023.100786

[5] H. Svensen. *Synthesis and Functionalization of 3-Nitropyridines*. Ph.D. thesis, Norwegian University of Science and Technology, Trondheim, Norwey (2001). http://hdl.handle.net/11250/244442

[6] Changgeng Qian, Tongchuan Li, T. Shen, L. Libertine-Garahan, J. Eckman, T. Biftu, S. Ip. Epibatidine is a nicotinic analgesic, *European Journal of Pharmacology* **250** (1993) R13-R14. https://doi.org/10.1016/0014-2999(93)90043-H

[7] A. Steiner, J. M. Mayer, B. Testa. Nicotinate esters: their binding to and hydrolysis by human serum albumin, *Journal of Pharmacy and Pharmacology* **44** (1992) 745-749. http://doi.org/10.1111/j.2042-7158.1992.tb05512.x

[8] J. S. Finch, T. J. DeKornfeld. Clonixin: A clinical evaluation of a new oral analgesic, *The Journal of Clinical Pharmacology and New Drugs* **11** (1971) 371-377. http://doi.org/10.1177/009127007101100508

[9] Z. Tilyabaev, A. A. Abduvakhabov. Alkaloids of anabasis aphylla and their cholinergic activities, *Chemistry of Natural Compounds* **34** (1998) 295-297. https://doi.org/10.1007/BF02282405

[10] A. S. Leal, K. Zydeck, S. Carapellucci, L. A. Reich, D. Zhang, J. A. Moerland, M. B. Sporn, K. T. Liby. Retinoid x receptor agonist lg100268 modulates the immune microenvironment in preclinical breast cancer models, *NPJ Breast Cancer* **5** (2019) 39. http://doi.org/10.1038/s41523-019-0135-5

[11] E. M. A. Antibacterial evaluation and molecular properties of pyrazolo[3,4-b]pyridines and thieno[2,3-b] pyridines, *Journal of Applied Pharmaceutical Science* **11** (2021) 118-124. https://doi.org/10.7324/JAPS.2021.110614

[12] T. J. Donohoe, C. R. Jones, A. F. Kornahrens, L. C. A. Barbosa, L. J. Walport, M. R. Tatton, M. O'Hagan, A. H. Rathi, D. B. Baker. Total synthesis of the antitumor antibiotic (±)-streptonigrin: First- and second-generation routes for de novo pyridine formation using ring-closing metathesis, *Journal of Organic Chemistry* **78** (2013) 12338-12350. https://doi.org/10.1021/jo402388f

[13] H. Hoehn, I. Polacek, E. Schulze. Potential antidiabetic agents. pyrazolo[3,4-b]pyridines, *Journal of Medicinal Chemistry* **16** (1973) 1340-1346. https://doi.org/10.1021/jm00270a006

[14] M. Croisy-Delcey, A. Croisy, D. Carrez, C. Huel, A. Chiaroni, P. Ducrot, E. Bisagni, L. Jin, G. Leclercq. Diphenyl quinolines and isoquinolines: synthesis and primary biological evaluation, *Bioorganic & Medicinal Chemistry* **8** (2000) 2629-2641. https://doi.org/10.1016/S0968-0896(00)00194-2

[15] A. Özdemir, G. Turan-Zitouni, Z. Asım Kaplancıklı, G. İşcan, S. Khan, F. Demirci. Synthesis and the selective antifungal activity of 5,6,7,8-tetrahydroimidazo[1,2-a]pyridine derivatives, *European Journal of Medicinal Chemistry* **45** (2010) 2080-2084. https://doi.org/10.1016/j.ejmech.2009.12.023

[16] M. Koudad, S. Dadou, F. Abrigach, A. E. Aatiaoui, M. Azzouzi, A. Oussaid, N. Benchat, M. Allali, K. Karrouchi. Synthesis and antimicrobial, antioxidant, ADME-T, and molecular docking studies of imidazo[1,2-a]pyridine derivatives, *Russian Journal of Organic Chemistry* **59** (2023) 1237-1247. https://doi.org/10.1134/S1070428023070163

[17] P. Makam, R. Kankanala, A. Prakash, T. Kannan. 2-(2-hydrazinyl)thiazole derivatives: Design, synthesis and in vitro antimycobacterial studies, *European Journal of Medicinal Chemistry* **69** (2013) 564-576. https://doi.org/10.1016/j.ejmech.2013.08.054

[18] V. Reddy, A. S. Jadhav, R. Vijaya Anand. A room-temperature protocol to access isoquinolines through ag(i) catalysed annulation of o-(1-alkynyl)arylaldehydes and ketones with nh4oac: elaboration to berberine and palmatine, *Organic and Biomolecular Chemistry* **13** (2015) 3732-3741. http://dx.doi.org/10.1039/C4OB02641A

[19] I. Boček Pavlinac, L. Persoons, D. Daelemans, K. Starčević, R. Vianello, M. Hranjec. Novel acrylonitrile derived imidazo[4,5-b]pyridines as antioxidants and potent antiproliferative agents for pancreatic adenocarcinoma, *International Journal of Biological Macromolecules* **266** (2024) 131239. https://doi.org/10.1016/j.ijbiomac.2024.131239

[20] S. Marın, R. Ibarra, M. Medina, P. Jansen. Sensitivity of caligus rogercresseyi (boxshall and bravo 2000) to pyrethroids and azamethiphos measured using bioassay tests—a large scale spatial study, *Preventive Veterinary Medicine* **122** (2015) 33-41. https://doi.org/10.1016/j.prevetmed.2015.09.017

[21] M. Ferrari. Effects of papaverine on smooth muscle and their mechanisms, *Pharmacological Research Communications* **6** (1974) 97-115. https://doi.org/10.1016/S0031-6989(74)80018-4

[22] M. Imran. Ethionamide and prothionamide based coumarinyl-thiazole derivatives: Synthesis, antitubercular activity, toxicity investigations and molecular docking studies, *Pharmaceutical Chemistry Journal* **56** (2022) 1215-1225. https://doi.org/10.1007/ s11094-022-02782-0

[23] J. Zhang, J. Dai, X. Lan, Y. Zhao, F. Yang, H. Zhang, S. Tang, G. Liang, X. Wang, Q. Tang. Synthesis, bioevaluation and molecular dynamics of pyrrolo-pyridine benzamide derivatives as potential antitumor agents in vitro and in vivo, *European Journal of Medicinal Chemistry* **233** (2022) 114215. https://doi.org/10.1016/j.ejmech.2022.114215

[24] R. M. Kassab, M. H. Ibrahim, A. Rushdi, S. J. Almehmadi, M. E. Zaki, S. A. Al-Hussain, Z. A. Muhammad, T. A. Farghaly. Comprehensive study for synthesis, antiviral activity, docking and ADME study for the new fluorinated hydrazonal and indeno[1,2- b]pyridine derivatives, *J. Mol. Struct.* **1305** (2024) 137752. https://doi.org/10.1016/ j.molstruc.2024.137752

[25] B. Turovska, I. Goba, A. Lielpetere, V. Glezer. Electrochemistry of pyridine derivatives, *Journal of Solid State Electrochemistry* **27** (2023) 1717-1729. https://doi.org/10.1007/ s10008-023-05425-w

[26] L. Jäntschi. Modelling of acids and bases revisited, *Studia Universitatis Babes-Bolyai, Seria Chemia* **67** (2022) 73-92. https://doi.org/10.24193/subbchem.2022.4.05

[27] O. Ciocirlan, E.-M. Ungureanu, A.-A. Vasile (Corbei), A. Stefaniu. Properties assessment by quantum mechanical calculations for azulenes substituted with thiophen- or furan-vinyl-pyridine, *Symmetry* **14** (2022) 354. https://doi.org/10.3390/ sym14020354

[28] M. V. Diudea, I. Gutman, L. Jäntschi, *Molecular Topology,* Nova Science Publishers, Huntington, N.Y., 2001, p. 197-232. http://lccn.loc.gov/2001031282

[29] L. Jäntschi. MDF - a new qsar/qspr molecular descriptors family, *Leonardo Journal of Sciences* **3** (2004) 68-85. http://ljs.academicdirect.org/A04/68_85.pdf

[30] S. D. Bolboacă, L. Jäntschi. Predictivity approach for quantitative structure-property models. Application for blood-brain barrier permeation of diverse drug-like compounds, *International Journal of Molecular Sciences* **12** (2011) 4348-4364. https://doi.org/10.3390/ijms12074348

[31] R. E. Sestras, , L. Jäntschi, S. D. Bolboacă. Quantum mechanics study on a series of steroids relating separation with structure, *JPC - Journal of Planar Chromatography - Modern TLC* **25** (2012) 528-533. https://doi.org/10.1556/JPC.25.2012.6.7

[32] S. D. Bolboacă, L. Jäntschi. Nanoquantitative structure-property relationship modeling on C_{42} fullerene isomers, *Journal of Chemistry* **2016** (2016) 1791756. https://doi.org/10.1155/2016/1791756

[33] L. Jäntschi, S. D. Bolboacă. *Chapter 15: Families of molecular descriptors*, in *New Frontiers in Nanochemistry: Concepts, Theories, and Trends. Volume 1: Structural Nanochemistry*, M. V. Putz, Ed., Apple Academic Press, Burlington, Canada, 2020, p. 143-169. https://doi.org/10.1201/9780429022937

[34] D. M. Opris, , M. V. Diudea. Peptide property modeling by cluj indices, *SAR and QSAR in Environmental Research* **12** (2001) 159-179. https://doi.org/10.1080/10629360108035377
[35] L. Jäntschi, S.-D. Bolboacă. Results from the use of molecular descriptors family on structure property/activity relationships, *International Journal of Molecular Sciences* **8** (2007) 189-203. https://doi.org/10.3390/i8030189
[36] D. Janežič, L. Jäntschi, D. S. Bolboacă. Sugars and sweeteners: Structure, properties and in silico modeling, *Current Medicinal Chemistry* **27** (2020) 5-22. https://doi.org/10.2174/0929867325666180926144401
[37] D. Bálint, L. Jäntschi. Comparison of molecular geometry optimization methods based on molecular descriptors, *Mathematics* **9** (2021) 2855. https://doi.org/10.3390/ math9222855
[38] L. Jäntschi. *Structure vs. Property: Models and Algorithms* (Babeș-Bolyai University: Habilitation Thesis) (2012). http://lori.academicdirect.org/doctoral/advisor/HabilThesis_InShort.pdf
[39] L. Jäntschi. *General Chemistry* (3rd ed.), AcademicDirect, Cluj-Napoca, Romania, 2013, pp. 436. http://ph.academicdirect.org/GCC_v3.pdf
[40] L. Jäntschi. Eigenproblem basics and algorithms, *Symmetry* **15** (2023) 2046. https://doi.org/10.3390/sym15112046
[41] E. Hückel. Quantentheoretische beiträge zum benzolproblem, *Zeitschrift für Physik* **70** (1931) 204-286. https://doi.org/10.1007/BF01339530
[42] D.-M. Joița, L. Jäntschi. Extending the characteristic polynomial for characterization of C_{20} fullerene congeners, *Mathematics* **5** (2017) 84. http://doi.org/10.3390/ math5040084
[43] S. D. Bolboacă, L. Jäntschi. How good can the characteristic polynomial be for correlations?, *International Journal of Molecular Sciences* **8** (2007) 335-345. https://doi.org/10.3390/i8040335
[44] L. Jäntschi. Structure-property relationships for solubility of monosaccharides, *Applied Water Science* **9** (2019) 38. https://doi.org/10.1007/s13201-019-0912-1
[45] T. A. Halgren. Merck molecular force field. II. MMFF94 van der waals and electrostatic parameters for intermolecular interactions, *Journal of Computational Chemistry* **17** (1996) 520-552. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<520::AID-JCC2>3.0.CO;2-W
[46] S. E. Brownell, S. Freeman, M. P. Wenderoth, A. J. Crowe. Biocore guide: A tool for interpreting the core concepts of vision and change for biology majors, *CBE-Life Sciences Education* **13** (2014) 200-211. https://doi.org/10.1187/cbe.13-12-0233
[47] A. C. Brown, T. R. Fraser. V.—On the connection between chemical constitution and physiological action. Part. i.-On the physiological action of the salts of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia, and nicotia, *Transactions of the Royal Society of Edinburgh* **25** (1868) 151-203
[48] R. S. DeFever, H. Bruce, G. Bhattacharyya. Mental rolodexing: Senior chemistry majors' understanding of chemical and physical properties, *Journal of Chemical Education* **92** (2015) 415-426. https://doi.org/10.1021/ed500360g
[49] L. Jäntschi. Detecting extreme values with order statistics in samples from continuous distributions, *Mathematics* **8** (2020) 216. https://doi.org/10.3390/math8020216
[50] L. Jäntschi. A test detecting the outliers for continuous distributions based on the cumulative distribution function of the data being tested, *Symmetry* **11** (2019) 835. https://doi.org/10.3390/sym11060835
[51] L. L. Pruteanu, L. Jäntschi, M. L. Unguresan, S. D. Bolboacă. Models of monovalent ions dissolved in water, *Studia Universitatis Babes-Bolyai, Seria Chemia* **61** (2016) 151-163. https://chem.ubbcluj.ro/~studiachemia/issues/chemia2016_1/15Pruteanu_etal_151_162.pdf

[52] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dear- den, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard, A. Tropsha. Qsar modeling: Where have you been? Where are you going to?, *Journal of Medicinal Chemistry* **57** (2014) 4977-5010. https://doi.org/10.1021/jm4004285

[53] J. Verma, V. M. Khedkar, E. C. Coutinho. 3d-qsar in drug design - a review, *Current Topics in Medicinal Chemistry* **10** (2010) 95-115. https://doi.org/10.2174/156802610790232260

[54] J. Huuskonen. Estimation of aqueous solubility for a diverse set of organic compounds based on molecular topology, *Journal of Chemical Information and Computer Sciences* **40** (2000) 773-777. https://doi.org/10.1021/ci9901338

[55] T. A. Halgren. Merck molecular force field. i. Basis, form, scope, parame- terization, and performance of MMFF94, *Journal of Computational Chemistry* **17** (1996) 490-519. https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6<490::AID-JCC1>3.0.CO;2-P

[56] L. Jäntschi. Triple crossed 3 C_{26} cyclic cumulene catenane, *Fullerenes, Nanotubes and Carbon Nanostructures* **32** (2024) 1-15. https://doi.org/10.1080/1536383X.2024.2354721

[57] S. N. Roese, J. D. Heintz, C. B. Uzat, A. J. Schmidt, G. V. Margulis, S. J. Sabatino, A. S. Paluch. Assessment of the sm12, sm8, and smd solvation models for predicting limiting activity coefficients at 298.15 K, *Processes* **8** (2020) 623. https://doi.org/10.3390/ pr8050623

[58] R. A. Fisher. On an absolute criterion for fitting frequency curves, *Messenger of Mathematics* **41** (1912) 155-160. https://hdl.handle.net/2440/15165

[59] D. G. Jenkins, P. F. Quintana-Ascencio. A solution to minimum sample size for regressions, *PloS One* **15** (2020) e0229345. https://doi.org/10.1371/journal.pone. 0229345

[60] K. Kelley, S. E. Maxwell. Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant, *Psychological Methods* **8** (2003) 305-321. https://doi.org/10.1037/1082-989X.8.3.305

[61] L. Jäntschi. Distribution fitting 1. Parameters estimation under assumption of agreement between observation and model, *Bulletin of University of Agricultural Sciences and Veterinary Medicine Cluj-Napoca. Horticulture* **66** (2009) 684-690. https://doi.org/10.48550/arXiv.0907.2829

[62] J. H. Steiger. Tests for comparing elements of a correlation matrix, *Psychological Bulletin* **87** (1980) 245-251. https://doi.org/10.1037/0033-2909.87.2.245

[63] L. Jäntschi. The eigenproblem translated for alignment of molecules, *Symmetry* **11** (2019) 1027. https://doi.org/10.3390/sym11081027

[64] K. Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, *The Philosophical Magazine series 5* **50** (1900) 157-175. https://doi.org/10.1080/14786440009463897

[65] W. S. Gosset. The probable error of a mean, *Biometrika* **6** (1908) 1-25. https://doi.org/10.1093/biomet/6.1.1

[66] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange, *The Journal of Chemical Physics* **98** (1993) 5648-5652. http://doi.org/10.1063/1.464913

[67] P. J. Stephens, F. J. Devlin, C. S. Ashvar, C. F. Chabalowski, M. J. Frisch. Theoretical calculation of vibrational circular dichroism spectra, *Faraday Discussions* **99** (1994) 103-119. http://doi.org/10.1039/FD9949900103

[68] Y. Zhao, D. G. Truhlar. The M06 suite of density functionals for main group thermo-chemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals, *Theoretical Chemistry Accounts* **120** (2008) 215-241. https://doi.org/10.1007/s00214-007-0310-x

[69] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, D. G. Truhlar. Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges, *Journal of Chemical Theory and Computation* **3** (2007) 2011-2033. https://doi.org/10.1021/ct7001418

[70] R. A. Fisher. On a distribution yielding the error functions of several well known statistics, *Proceedings of the International Mathematical Congress. Toronto* **2** (1924) 805-813. https://hdl.handle.net/2440/ 15183