Inter*net* Electronic Journal of Molecular Design

October 2002, Volume 1, Number 10, Pages 559–571

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday
Part 2

Guest Editor: Jun–ichi Aihara

QSAR Models for the Dermal Penetration of Polycyclic Aromatic Hydrocarbons

Ovidiu Ivanciuc,[1] Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University "Politehnica" of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Received: June 17, 2002; Revised: August 27, 2002; Accepted: September 21, 2002; Published: October 31, 2002

Citation of the article:
O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, QSAR Models for the Dermal Penetration of Polycyclic Aromatic Hydrocarbons, *Internet Electron. J. Mol. Des.* **2002**, *1*, 559–571, http://www.biochempress.com.

Copyright © 2002 BioChem Press

Internet Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

QSAR Models for the Dermal Penetration of Polycyclic Aromatic Hydrocarbons[#]

Ovidiu Ivanciuc,[1,*] Teodora Ivanciuc,[2] and Alexandru T. Balaban[3]

[1] Sealy Center for Structural Biology, Department of Human Biological Chemistry & Genetics, University of Texas Medical Branch, Galveston, Texas 77555–1157
[2] University Politehnica of Bucharest, Department of Organic Chemistry, Bucharest, Romania
[3] Texas A&M University at Galveston, Department of Oceanography and Marine Sciences, 5007 Avenue U, Galveston, Texas 77551

Received: June 17, 2002; Revised: August 27, 2002; Accepted: September 21, 2002; Published: October 31, 2002

Internet Electron. J. Mol. Des. **2002**, *1* (10), 559–571

Abstract

Quantitative structure–activity relationships (QSAR) models for the dermal penetration of 60 polycyclic aromatic hydrocarbons (PAHs) were established with the CODESSA program. The QSAR models for the PAH dermal penetration are obtained by selecting descriptors from a wide diversity of constitutional, topological, electrostatic and quantum structural indices. Standard quantum chemistry packages are used for optimizing the molecular geometry and for semi–empirical quantum computations. A heuristic algorithm selects the best multiple linear regression equation according to the highest statistical indices; the predictive power of each QSAR is estimated with the leave–one–out (LOO) cross–validation method. The best QSAR model with two descriptors has r^2 = 0.748, r^2_{LOO} = 0.719, s = 6.5, and F = 84.42. A significant improvement of the statistical indices is obtained by adding a third theoretical descriptor, *i.e.* r^2 = 0.761, r^2_{LOO} = 0.725, s = 6.4, and F = 59.39. Our results demonstrate that QSAR models can be used in risk assessment studies in order to estimate the dermal penetration properties of PAHs from mineral oils, coal, tar, and derived products.

Keywords. Quantitative structure–activity relationships; QSAR; dermal penetration; polycyclic aromatic hydrocarbon.

1 INTRODUCTION

During the last twenty years quantitative structure–property relationships (QSPR) and quantitative structure–activity relationships (QSAR) models have gained an extensive recognition in physical, organic, analytical, pharmaceutical and medicinal chemistry, biochemistry, chemical engineering and technology, toxicology, and environmental sciences. The main contributions to the widespread use of QSPR and QSAR models come from the development of novel structural descriptors and statistical equations relating various physical, chemical, and biological properties to

[#] Dedicated to Professor Haruo Hosoya on the occasion of the 65th birthday.

* Correspondence author; E–mail: ivanciuc@netscape.net.

the chemical structure. The success of the QSPR and QSAR approach can be explained by the insight offered into the structural determination of chemical properties, and the possibility to estimate the properties of new chemical compounds without the need to synthesize and test them. The main hypothesis in the QSPR and QSAR approach is that all properties (physical, chemical, and biological) of a chemical substance are statistically related to its molecular structure.

The dermal penetration of drugs or toxic compounds is usually investigated (*a*) by monitoring their *in vivo* release in human volunteers or live animals, (*b*) by using excised skin from human or animal sources, or (*c*) with experiments employing synthetic membranes that model the skin diffusion of chemicals [1–4]. The amount of a chemical which is able to penetrate the skin, together with its concentration–dependent distribution within the skin, are important factors in the development of formulations for topical application and transdermal delivery, and for risk assessment of xenobiotics. In order to reach the systemic circulation, any chemical that comes into contact with the skin has to pass through both the *stratum corneum* and the viable tissue. Whenever a chemical has a potential toxic, mutagenic, or carcinogenic action, its *in vivo* testing in humans is substituted with *in vitro* methods. The skin contact with polycyclic aromatic hydrocarbons (PAHs) represents a potential health risk because several PAHs found in mineral oils, coal–tars, and derived products have been shown to be mutagenic and carcinogenic. Dermal exposure to PAHs can occur by deposition of vapors and particles, by splashing of oils, or by contact with contaminated soil. Recently, Roy and co–workers determined the dermal penetration of 60 PAHs using skin excised from rats [5]. From an initial set of more than 50 structural descriptors, 18 were selected to generate multiple linear regression (MLR) models. With the same experimental database Gute *et al*. used a hierarchical QSAR approach to obtain several reliable QSAR models for PAH dermal penetration [6]. An artificial neural network (ANN) structure–dermal penetration model was proposed by Devillers for the same 60 compounds, with the intention to explore the nonlinear relationships between the structural descriptors and skin penetration properties of PAHs [7]. Another QSAR model for the dermal penetration for this set of PAHs was developed using molecular quantum similarity measures [8]. In this paper we improve the dermal penetration QSAR models for PAHs by selecting the structural descriptors for the multilinear regression equation from a wide range of topological, geometrical, electrostatic, and quantum indices.

2 MATERIALS AND METHODS

The investigation of large and diverse molecular databases was made possible by the advent of general QSPR/QSAR programs [9,10], such as CODESSA [11–15], which integrate the computation of structural descriptors with the generation of structure–property models. These programs compute more than one thousand structural descriptors from five classes: constitutional, graph theoretic and topological indices, geometrical, electrostatic, and quantum–chemical

descriptors. Using statistical methods, such as MLR, PCA (principal components analysis), PLS (partial least squares), or ANN, the best descriptors are selected in the final QSAR model.

PAH Dermal Penetration Data. The experimental values for the dermal penetration of 60 PAHs (Figure 1) taken from the literature [5] are presented in Table 1, where PADA represents the percent of applied dose dermally absorbed over 24 hours. Dorsal skin from female Sprague–Dawley rats was lightly shaved with an electric clipper before excision, and skin sections (~350 μm) were prepared. The dermal penetration was determined by GC/FID and GC/MS.

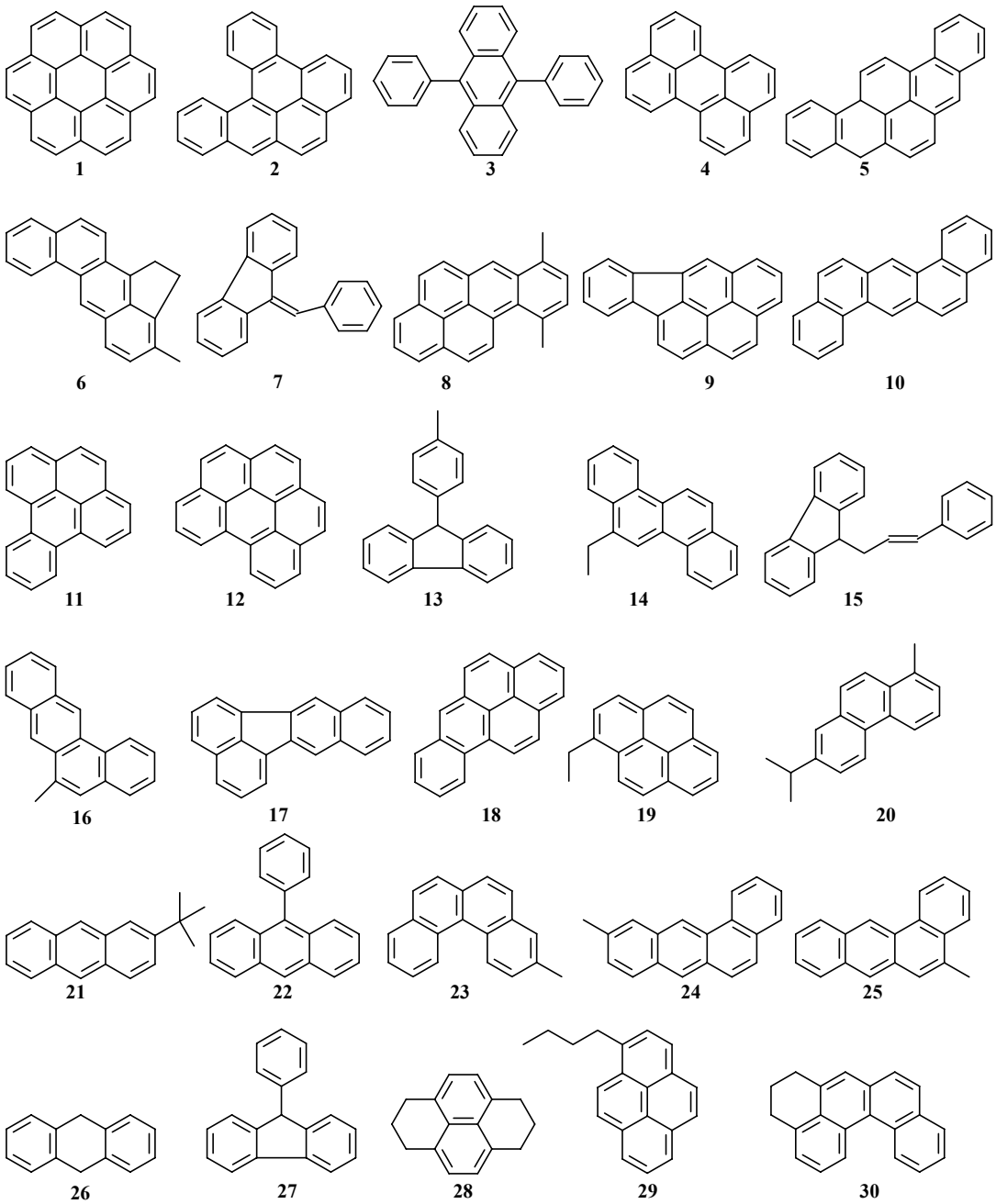

Figure 1. The structure of the 60 polycyclic aromatic hydrocarbons from Ref. [5].

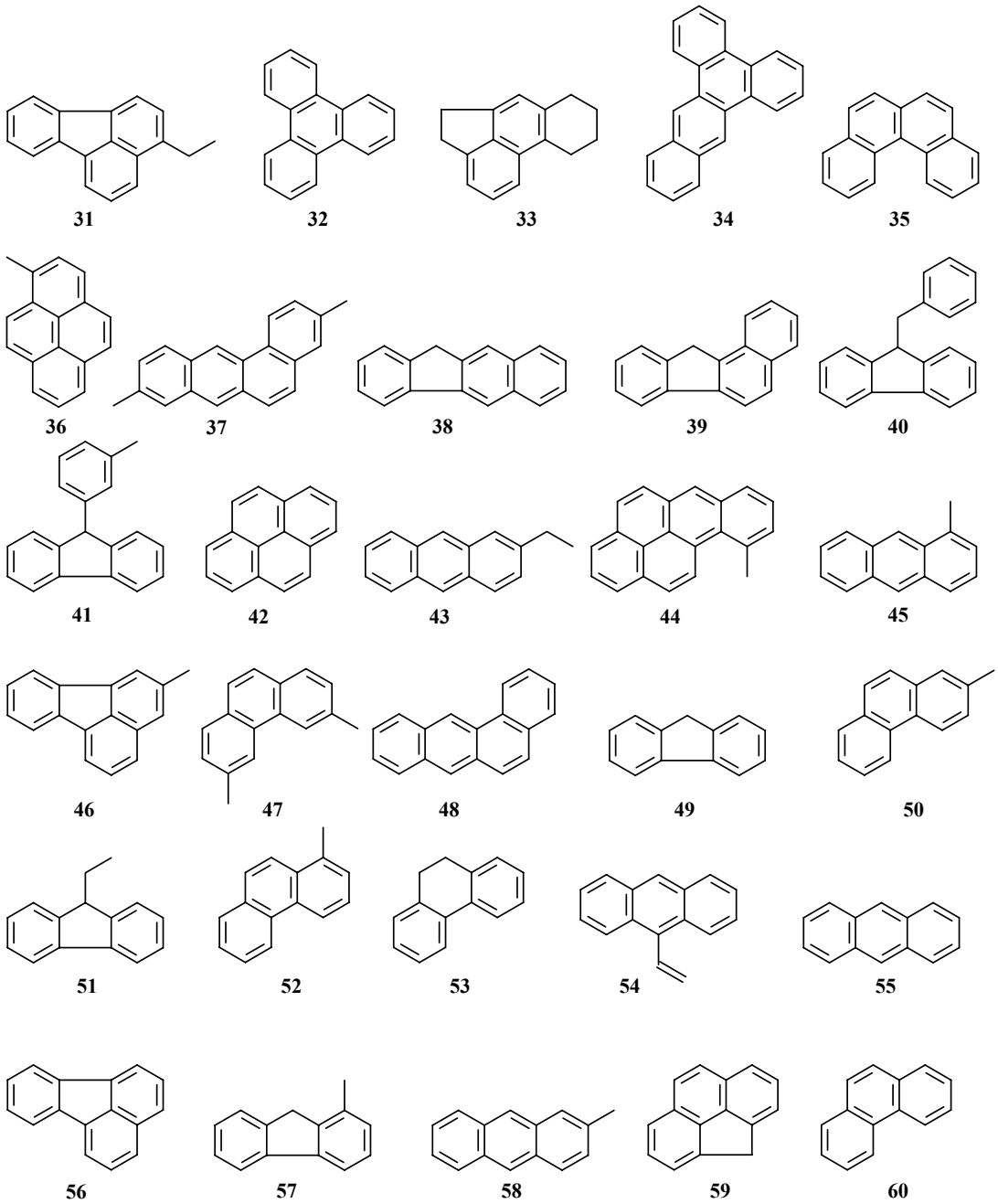

Figure 1. (Continued).

Previous QSAR Models. For the whole set of 60 PAHs, Roy *et al*. [5] obtained the following QSAR equation:

$$\text{PADA} = 111.9 - 14.7 \log P - 22.0\ \text{SHDW6} \qquad (1)$$
$$n = 60 \quad r^2 = 0.640 \quad s = 7.7 \quad \text{F} = 54$$

where log P is the calculated octanol–water partition coefficient, and SHDW6 is the normalized area of the two–dimensional projection of the molecule onto the Y–Z plane. For a subset of 22 PAHs (non–hydro and substituted) the QSAR model contains only the log P descriptor [5]:

$$\text{PADA} = 121.4 - 19.6 \log P \qquad n = 22 \quad r^2 = 0.792 \quad s = 7.2 \quad F = 73 \tag{2}$$

For the subset of 38 non–hydro, unsubstituted, or methyl–substituted PAHs, a good correlation is obtained with log *P* and EDEN, the most negative value for the approximate atomic σ–electron density calculated as $(Q_i - NBD_i)/VDW_i$, where Q_i is the partial atomic charge of atom *i*, NBD_i is the number of nonbonded electrons of atom *i*, and VDW_i is the van der Waals radius of atom *i* [5]:

$$\text{PADA} = 104.7 - 16.9 \log P - 1405\ \text{EDEN} \qquad n = 38 \quad r^2 = 0.689 \quad s = 8.0 \quad F = 41 \tag{3}$$

Table 1. The Name of the 60 PAHs from Figure 1, their Experimental, Computed and Residual PADA Obtained with Eq. (10).

PAH	Compound	PADA		
		exp.	cal.	res.
1	coronene	0.7	–0.1	0.8
2	dibenzo[*a*,*l*]pyrene	2.0	5.3	–3.3
3	9,10–diphenylanthracene	6.0	5.4	0.6
4	perylene	7.0	16.4	–9.4
5	dibenzo[*a*,*i*]pyrene	8.0	4.8	3.2
6	3–methylcholanthene	8.0	12.9	–4.9
7	9–benzylidenefluorene	8.0	18.9	–10.9
8	7,10–dimethylbenzo[*a*]pyrene	8.3	11.3	–3.0
9	indeno[1,2,3:*c*,*d*]pyrene	9.0	11.7	–2.7
10	dibenz[*a*,*h*]anthracene	9.4	9.4	0.0
11	benzo[*e*]pyrene	10.0	14.0	–4.0
12	benzo[*g*,*h*,*i*]perylene	10.0	7.5	2.5
13	9–*p*–tolylfluorene	10.0	18.7	–8.7
14	6–ethylchrysene	10.0	16.9	–6.9
15	9–*E*–cinnamylfluorene	11.0	13.7	–2.7
16	6–methylbenz[*a*]anthracene	14.0	20.7	–6.7
17	benzo[*k*]fluoranthene	14.0	17.2	–3.2
18	benzo[*a*]pyrene	15.0	15.9	–0.9
19	1–ethylpyrene	18.0	23.7	–5.7
20	7–isopropyl–1–methylphenanthrene	20.0	20.5	–0.5
21	2–*tert*–butylanthracene	20.0	26.3	–6.3
22	9–phenylanthracene	20.0	19.3	0.7
23	3–methylbenzo[*c*]phenanthrene	20.0	20.9	–0.9
24	10–methylbenz[*a*]anthracene	20.0	21.0	–1.0
25	5–methylbenz[*a*]anthracene	20.0	20.8	–0.8
26	9,10–dihydroanthracene	20.0	29.4	–9.4
27	9–phenylfluorene	20.0	20.8	–0.8
28	1,2,3,6,7,8–hexahydropyrene	20.0	20.6	–0.6
29	*n*–butylpyrene	20.0	16.1	3.9
30	5,6–dihydro–4*H*–dibenz[*a*,*k*,*l*]anthracene	20.0	11.2	8.8
31	3–ethylfluoranthene	20.0	27.2	–7.2
32	triphenylene	20.0	20.6	–0.6
33	7,8,9,10–tetrahydroacephenanthrene	20.0	23.3	–3.3
34	2,3–benztriphenylene	20.0	9.3	10.7
35	benzo[*c*]phenanthrene	20.0	23.4	–3.4
36	1–methylpyrene	22.0	27.1	–5.1
37	3,9–dimethylbenz[*a*]anthracene	24.0	16.0	8.0
38	2,3–benzofluorene	25.0	26.8	–1.8
39	1,2–benzofluorene	25.0	25.8	–0.8
40	9–benzylfluorene	26.0	19.6	6.4
41	9–*m*–tolylfluorene	29.0	18.9	10.1

Table 1. (Continued)

PAH	Compound	PADA		
		exp.	cal.	res.
42	pyrene	30.0	31.2	–1.2
43	2–ethylanthracene	30.0	34.7	–4.7
44	10–methylbenzo[*a*]pyrene	32.0	15.2	16.8
45	1–methylanthracene	32.0	39.9	–7.9
46	2–methylfluoranthene	33.0	31.5	1.5
47	3,6–dimethylphenanthrene	33.0	28.6	4.4
48	benz[*a*]anthracene	35.0	23.5	11.5
49	fluorene	36.0	47.4	–11.4
50	2–methylphenanthrene	36.0	36.6	–0.6
51	9–ethylfluorene	38.0	37.6	0.4
52	1–methylphenanthrene	40.0	35.8	4.2
53	9,10–dihydrophenanthrene	40.0	39.3	0.7
54	9–vinylanthracene	40.0	35.9	4.1
55	anthracene	42.0	45.7	–3.7
56	fluoranthene	42.0	35.2	6.8
57	1–methylfluorene	49.0	41.3	7.7
58	2–methylanthracene	50.0	40.2	9.8
59	4*H*–cyclopenta[*d*,*e*,*f*]phenanthrene	50.0	37.0	13.0
60	phenanthrene	50.0	41.5	8.5

Gute *et al*. improved the QSAR models from Eq. (1) by using structural descriptors computed with POLLY [6]. The use of the number of paths of length 0, P_0, significantly improved the QSAR:

$$\text{PADA} = 224.1 - 67.9\, P_0 \qquad n = 60 \quad r^2 = 0.675 \quad s = 7.4 \quad \text{F} = 120.6 \tag{4}$$

The use of the topological index ${}^1\chi^b$, the bond path connectivity index of order 1, further improved the correlation of the dermal penetration with the PAH structure [6]:

$$\text{PADA} = 179.7 - 78.8\, {}^1\chi^b \qquad n = 60 \quad r^2 = 0.695 \quad s = 7.1 \quad \text{F} = 132.0 \tag{5}$$

The QSAR with ${}^{3D}W$, the three–dimensional Wiener index computed from the geometric (Euclidean) distance matrix of the hydrogen–suppressed molecule, was almost as good as the one obtained with P_0 [6]:

$$\text{PADA} = 186.0 - 25.4\, {}^{3D}W \qquad n = 60 \quad r^2 = 0.673 \quad s = 7.4 \quad \text{F} = 119.3 \tag{6}$$

However, P_0 was computed from the molecular graph while ${}^{3D}W$ was computed from the molecular geometry, making P_0 a much more attractive descriptor for modeling the PAH dermal penetration. A fairly good QSAR model was obtained by using the molecular weight *MW* [6]:

$$\text{PADA} = 90.6 - 0.3\, MW \qquad n = 60 \quad r^2 = 0.674 \quad s = 7.4 \quad \text{F} = 120.0 \tag{7}$$

Various artificial neural network (ANN) models were investigated by Devillers for the same 60 compounds, with the intention to explore the nonlinear relationships between the structural descriptors and skin penetration properties of PAHs [7]. The best results were obtained with a

multiplayer perceptron ANN with 6 input neurons, 3 hidden neurons, and one output neuron. The 6 input structural descriptors were log *P*, *MW*, $^1\chi^b$, SHDW6, an indicator variable for the presence of hydro–PAH, and an indicator variable for PAHs with 5 or more cycles. However, the ANN model contains 25 optimizable parameters (connections between neurons) while Eqs. (4)–(7) contain each 2 optimizable parameters.

Molecular quantum similarity measures computed at the 3–21G level were also used to model the dermal penetration of the 60 PAHs [8]. The structural information from the quantum similarity matrix for the 60 PAHs was analyzed with PCA (principal components analysis) and the first 17 PCs (principal components) were retained for developing QSAR models. The best QSAR model (r^2 = 0.684 and r^2_{LOO} = 0.634) was obtained with three PCs (namely 1, 2, and 13).

Molecular Modeling. In the present investigation, the chemical structures were generated with HyperChem [16], the geometry optimization was performed with MOPAC [17] using the semiempirical quantum method AM1 [18] and the QSAR models were computed with CODESSA [19].

Structural Descriptors. The HyperChem structure files and the MOPAC output files were used by the CODESSA program to calculate 328 descriptors for the 60 PAHs. CODESSA computes five classes of structural descriptors: constitutional (number of various types of atoms and bonds, number of rings, molecular weight); topological (Wiener index, Randić connectivity indices, Kier shape indices, information theory indices); geometrical (principal moments of inertia, shadow indices, molecular volume and surface area); electrostatic (when atomic charges are computed on the basis of atomic electronegativity: minimum and maximum partial charges, polarity parameter, charged partial surface area descriptors, hydrogen bond donor and acceptor surface indices); quantum-chemical (minimum and maximum partial charges, Fukui reactivity indices, dipole moment, HOMO and LUMO energies, molecular polarizability, minimum/maximum valency of an atom, minimum/maximum electron–electron repulsion for an atom, minimum/maximum exchange energy for a chemical bond, minimum/maximum atomic orbital electronic population, minimum/maximum nucleus–nucleus repulsion for a chemical bond, minimum/maximum electron–nucleus attraction for a chemical bond).

Multiple Linear Regression Model. From the whole set of 328 descriptors generated with CODESSA we have discarded descriptors with a constant value for all molecules in the data set. Descriptors for which values were not available for every molecule were assigned a zero value for the missing position. In the next step the number of descriptors was reduced by eliminating those with F–test values less than 1, t–test values less than 0.1 or correlation coefficients with the dermal penetration less than 0.1; as a result of this descriptor selection procedure, 204 descriptors remained for 60 PAHs. CODESSA develops MLR models by a heuristic method which includes the following steps: (*a*) All quasi–orthogonal pairs of structural descriptors are selected from the initial

set. Two descriptors are considered orthogonal if their intercorrelation coefficient r_{ij} is lower than 0.1. (*b*) CODESSA uses the pairs of orthogonal descriptors to compute the biparametric regression equations. The most significant 10 pairs of molecular descriptors are used in the third step. (*c*) To an MLR model containing *n* descriptors a new descriptor is added to generate a model with *n*+1 descriptors if the new descriptor is not significantly correlated with the previous *n* descriptors (intercorrelation coefficient lower than 0.8). Step (*c*) is repeated until MLR models with a prescribed number of descriptors are obtained.

Table 2. Notation of the Structural Descriptors Involved in the QSAR Models for the Dermal Penetration of Polycyclic Aromatic Compounds.

Notation	Descriptor
SD1	average electrophilic reactivity index for a C atom
SD2	${}^{1}\chi^{v}$, Kier and Hall valence connectivity index of order 1
SD3	total molecular 2–center resonance energy
SD4	${}^{2}\chi^{v}$, Kier and Hall valence connectivity index of order 2
SD5	gravitation index for all pairs of bonded atoms
SD6	${}^{3}\chi^{v}$, Kier and Hall valence connectivity index of order 3
SD7	number of bonds
SD8	${}^{3}\chi$, Randić connectivity index of order 3
SD9	CIC_0, complementary information content of order 0
SD10	log *P*, the octanol/water partition coefficient taken from Ref. [5]
SD11	average valence of a C atom
SD12	XY shadow / XY rectangle
SD13	minimum electron–electron repulsion for a C–H bond
SD14	maximum nucleus–nucleus repulsion for a C–H bond
SD15	minimum (>0.1) bond order of a H atom
SD16	maximum resonance energy for a C–H bond
SD17	FPSA3 = PPSA3/TMSA, fractional PPSA3 (quantum)
SD18	maximum 1–electron reactivity index for a C atom
SD19	RPCG, relative positive charged surface area (quantum)
SD20	relative number of rings
SD21	maximum quantum bond order of a H atom
SD22	E_{HOMO-1}, energy of the HOMO – 1 molecular orbital
SD23	RNCG, relative negative charged surface area (electrostatic)
SD24	maximum valence of a H atom
SD25	Q_{min}, minimum partial atomic charge
SD26	$ABIC_2$, average bonding information content of order 2
SD27	$ACIC_2$, average complementary information content of order 2
SD28	BIC_2, bonding information content of order 2
SD29	IC_2, information content of order 2

Model Validation. QSAR correlations can be observed not only because a causal relationship exists between a set of descriptors and a property, but also due to statistical bias resulting from errors in determining structural descriptors, experimental errors in measuring the property, or even due to chance alone. Model validation techniques are needed in order to distinguish between true and random correlations and to estimate the predictive power of the model. Although the QSAR equations developed with CODESSA are obtained by selection of descriptors from a large pool, several descriptor selection techniques are used in order to minimize the possibility of chance correlations. In a first step, from the initial pool of descriptors, CODESSA eliminates descriptors as

indicated above, thus greatly reducing the dimensionality of the problem – that of finding a QSAR equation with a good predictive power. Then, as described in the previous section, a heuristic algorithm selects only quasi–orthogonal groups of descriptors that are tested for correlation with the boiling temperatures of carbonyl compounds. This selection algorithm ensures that the probability of obtaining a chance correlation is low, and maintains a reasonable searching time. Finally, the leave–one–out (LOO) cross–validation procedure is applied to each and every MLR equation in order to estimate the prediction power of dermal penetration QSAR.

3 RESULTS AND DISCUSSION

Table 2 presents the notation and a short description of the structural descriptors involved in the QSAR models reported in this investigation; more complete definitions of the descriptors can be found in the CODESSA manuals [19]. The statistical results obtained in the best ten monoparametric correlations are presented in Table 3. For each equation we report the descriptor involved and the statistical indices of the model, *i.e.* the correlation coefficient r, the leave–one–out cross–validation correlation coefficient r_{cv}, the standard deviation s, and the F–test value.

Table 3. Structural Descriptors and Statistical Indices (Calibration Correlation Coefficient r, Leave–One–Out Cross–Validation Correlation Coefficient r_{LOO}, Standard Deviation s, and Fisher Test F) in the Best Ten Monoparametric QSAR Models for the Dermal Penetration of the 60 PAH.

No.	SD	r^2	r^2_{LOO}	s	F
1	SD1	0.7112	0.6896	6.9	142.82
2	SD2	0.6970	0.6745	7.1	133.44
3	SD3	0.6748	0.6493	7.4	120.36
4	SD4	0.6681	0.6464	7.4	116.73
5	SD5	0.6664	0.6424	7.4	115.84
6	SD6	0.6631	0.6418	7.5	114.14
7	SD7	0.6562	0.6304	7.6	110.72
8	SD8	0.6517	0.6281	7.6	108.50
9	SD9	0.6342	0.6024	7.8	100.57
10	SD10	0.6187	0.5880	8.0	94.09

The best PADA model from Table 3 is obtained with SD1, the average electrophilic reactivity index for a carbon atom:

$$\begin{gathered} \text{PADA} = -\,50.10(\pm 6.26) + 12542(\pm 1050)\ \text{SD1} \\ n = 60 \quad r^2 = 0.711 \quad r^2_{LOO} = 0.690 \quad s = 6.9 \quad \text{F} = 142.82 \end{gathered} \tag{8}$$

The QSAR models obtained with SD1 represents a significant improvement over previous models from Eqs. (1) and (4)–(7). In the group of 10 QSAR monoparametric models from Table 3 one can find a constitutional descriptor (SD7, the number of bonds), log P, five topological indices (${}^1\chi^v$, ${}^2\chi^v$, ${}^3\chi^v$, ${}^3\chi$, and CIC_0), a geometrical index (SD5, the gravitation index for all pairs of bonded atoms), and two quantum indices (the average electrophilic reactivity index for a carbon atom and the total molecular 2–center resonance energy).

Table 4. Structural Descriptors and Statistical Indices in the Best Ten QSAR Models with Two Descriptors for the Dermal Penetration of the 60 PAH.

No.	SD_1	SD_2	r^2	r^2_{LOO}	s	F
1	SD1	SD11	0.7476	0.7189	6.5	84.42
2	SD1	SD12	0.7373	0.7079	6.7	79.98
3	SD1	SD13	0.7359	0.7028	6.7	79.40
4	SD1	SD14	0.7350	0.7042	6.7	79.05
5	SD1	SD15	0.7338	0.7021	6.7	78.57
6	SD1	SD16	0.7307	0.6986	6.8	77.31
7	SD1	SD17	0.7299	0.6994	6.8	77.00
8	SD2	SD18	0.7250	0.6952	6.8	75.15
9	SD19	SD20	0.7211	0.6920	6.9	73.68
10	SD3	SD18	0.7142	0.6831	7.0	71.23

Improved PADA QSAR models are obtained by using two descriptors, as can be seen from the models in Table 4, with r^2 between 0.7476 and 0.7142, and the leave–one–out cross–validation correlation coefficient r^2_{LOO} between 0.7189 and 0.6831. In the first 7 bi–parametric models, SD1 appears together with other geometric, electrostatic, or quantum descriptors. The best model from Table 4 is obtained with SD1 and SD11, the average valence of a carbon atom:

$$\begin{gathered} \text{PADA} = 8897(\pm 3119) + 12025(\pm 1006)\ \text{SD1} - 2269(\pm 791)\ \text{SD11} \\ n = 60 \quad r^2 = 0.748 \quad r^2_{LOO} = 0.719 \quad s = 6.5 \quad F = 84.42 \end{gathered} \tag{9}$$

Table 5. Structural Descriptors and Statistical Indices in the Best Ten QSAR Models with Three Descriptors for the Dermal Penetration of 60 PAH.

No.	SD_1	SD_2	SD_3	r^2	r^2_{LOO}	s	F
1	SD1	SD11	SD21	0.7608	0.7251	6.4	59.39
2	SD1	SD11	SD22	0.7593	0.7211	6.4	58.89
3	SD1	SD11	SD23	0.7589	0.7225	6.4	58.77
4	SD1	SD11	SD24	0.7574	0.7214	6.5	58.27
5	SD1	SD11	SD18	0.7572	0.7194	6.5	58.20
6	SD1	SD11	SD25	0.7570	0.7209	6.5	58.14
7	SD1	SD11	SD26	0.7562	0.7212	6.5	57.90
8	SD1	SD11	SD27	0.7561	0.7210	6.5	57.87
9	SD1	SD11	SD28	0.7560	0.7215	6.5	57.82
10	SD1	SD11	SD29	0.7559	0.7215	6.5	57.81

A further improvement of the QSAR models is obtained by using three descriptors, as can be seen from Table 5, where the PADA models have r^2 between 0.7608 and 0.7559, and the leave–one–out cross–validation correlation coefficient r^2_{LOO} between 0.7251 and 0.7215. All QSAR models from Table 5 are derived from the best bi–parametric model containing SD1 and SD11. The best PADA model is obtained with SD1, SD11, and SD21, the maximum quantum bond order of a H atom:

$$\begin{gathered} \text{PADA} = 9947(\pm 3121) + 11754(\pm 999)\ \text{SD1} - 2566(\pm 795)\ \text{SD11} + 126.2(\pm 71.7)\ \text{SD21} \\ n = 60 \quad r^2 = 0.761 \quad r^2_{LOO} = 0.725 \quad s = 6.4 \quad F = 59.39 \end{gathered} \tag{10}$$

A further increase in the number of structural descriptors in the PADA models does not significantly improve the prediction power of the QSAR models. In Figure 1 we present the

experimental vs. calculated PADA for the group of 60 PAHs, while in Figure 2 we display the calibration residuals computed with Eq. (10). Both these figures show that there is no special trend of the residuals and no clusters can be detected in the data.

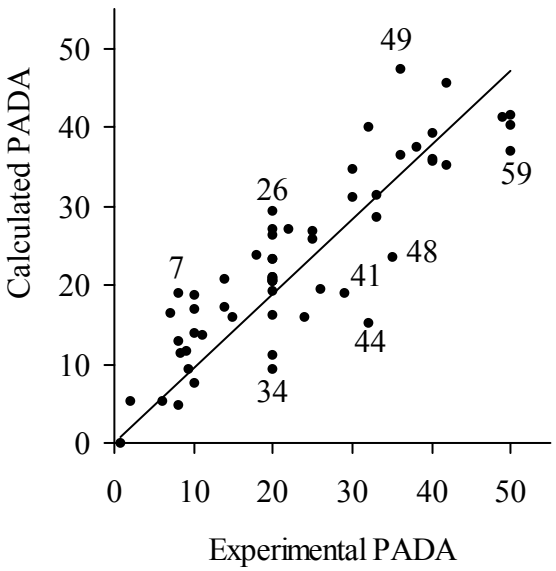

Figure 2. Experimental dermal penetration *vs* calculated with Eq. (10) for 60 PAHs.

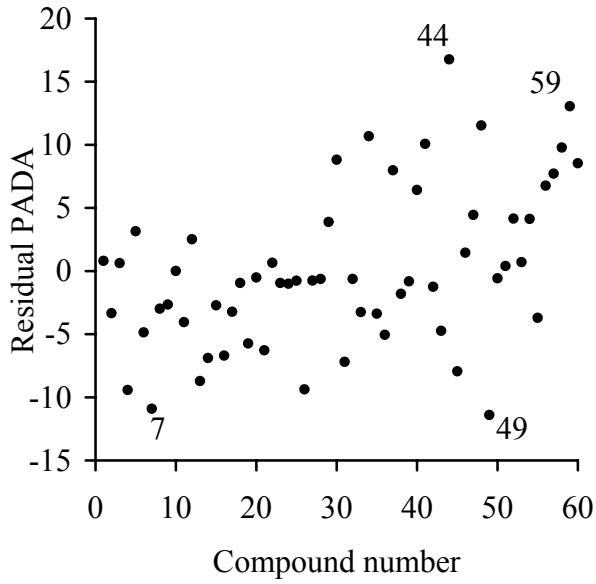

Figure 3. Calibration residuals computed with Eq. (10) for the dermal penetration of 60 PAHs.

Table 6. Distribution of the Absolute Values of the Residuals for QSAR Models from Eq. (1), Eq. (5), ANN, and Eq. (10).

Range	Eq. (1) Ref. [5]	Eq. (5) Ref. [6]	ANN Ref. [7]	Eq. (5)
<2	9	9	21(2)	19
[2, 4)	15	16	11(1)	11
[4, 6)	9	9	6(1)	8
[6, 8)	9	12	8	8
[8, 10)	7	6	5	7
[10, 12)	5	3	4(1)	5
[12, 14)	4	2	0	1
[14, 16)	0	1	0	0
≥16	2	2	0	1

The distribution of the absolute values of the residuals for QSAR models from Eq. (1) [5], Eq. (5) [6], ANN [7], and Eq. (10) is presented in Table 6. For the results obtained with the neural network, the residual values obtained for the testing set (PAH **9**, **18**, **27**, **36**, and **45**) are indicated in parentheses. In this comparison we have to consider that Eqs. (4)–(7) contain each 2 optimizable parameters, Eq. (10) has four optimizable parameters, while the ANN model contains 25 optimizable parameters. With a much lower number of optimizable parameters than the ANN, Eq. (10) has a distribution of the residuals very similar to the one obtained with the neural network, with the majority of absolute residuals lower than 2. Two compounds have residuals between 2*s* and 3*s*

(**59** with 13.0, and **44** with 16.8), and no statistical outlier (PAHs with absolute residual greater than 3*s*) is produced by Eq. (10).

4 CONCLUSIONS

A successful application of the CODESSA software system was presented in this study for the prediction of the dermal penetration of 60 polycyclic aromatic hydrocarbons using theoretical descriptors derived from the molecular structure. After a heuristic screening of relevant structural descriptors, the QSAR model with the best statistics (r^2 = 0.761, r^2_{LOO} = 0.725, s = 6.4, and F = 59.39) was obtained with three quantum descriptors: average electrophilic reactivity index for a C atom, average valence of a C atom, and maximum quantum bond order of a H atom. Together with the QSAR model proposed in this study, these three quantum descriptors could be used to estimate the dermal penetration for not yet synthesized or laboratory tested polycyclic aromatic hydrocarbons.

5 REFERENCES

[1] A. Wilschut, W. F. ten Berge, P. J. Robinson, and T. E. McKone, Estimating Skin Permeation. The Validation of Five Mathematical Skin Permeation Models, *Chemosphere* **1995**, *30*, 1275–1296.
[2] M. D. Barratt, Quantitative Structure–Activity Relationships for Skin Permeability, *Toxic. Vitro* **1995**, *9*, 27–37.
[3] F. Barbato, B. Cappello, A. Miro, M. I. La Rotonda, and F. Quaglia, Chromatographic Indexes on Immobilized Artificial Membranes for the Prediction of Transdermal Transport of Drugs, *Il Farmaco* **1998**, *53*, 655–661.
[4] K. Ohkura and H. Hori, Analysis of Structure–Permeability Correlation of Nitrophenol Analogues in Newborn Rat Abdominal Skin Using Semiempirical Molecular Orbital Calculation, *Bioorg. Med. Chem*. **1999**, *7*, 309–314.
[5] T. A. Roy, A. J. Krueger, C. R. Mackerer, W. Neil, A. M. Arroyo, and J. J. Yang, SAR Models for Estimating the Percutaneous Absorption of Polynuclear Aromatic Hydrocarbons, *SAR QSAR Environ. Res*. **1998**, *9*, 171–185.
[6] B. D. Gute, G. D. Grunwald, and S. C. Basak, Prediction of the Dermal Penetration of Polycyclic Aromatic Hydrocarbons (PAHs): A Hierarchical QSAR Approach, *SAR QSAR Environ. Res*. **1999**, *10*, 1–15.
[7] J. Devillers, Nonlinear Modeling of the Percutaneous Absorption of Polycyclic Aromatic Hydrocarbons, *Polycyc. Arom. Comp*. **2000**, *18*, 231–241.
[8] A. Gallegos, D. Robert, X. Gironés, and R. Carbó–Dorca, Structure–Toxicity Relationships of Polycyclic Aromatic Hydrocarbons using Molecular Quantum Similarity, *J. Comput.–Aided Mol. Des*. **2001**, *15*, 67–80.
[9] O. Ivanciuc and J. Devillers, Algorithms and Software for the Computation of Topological Indices and Structure–Property Models; in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. J. Devillers and A. T. Balaban, Gordon and Breach Science Publishers, Amsterdam, 1999, pp 779–804.
[10] A. R. Katritzky, U. Maran, V. S. Lobanov, and M. Karelson, Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 1–18.
[11] A. R. Katritzky, L. Mu, V. S. Lobanov, and M. Karelson, Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics, *J. Phys. Chem*. **1996**, *100*, 10400–10407.
[12] A. R. Katritzky, V. S. Lobanov, and M. Karelson, Normal Boiling Points for Organic Compounds: Correlation and Prediction by a Quantitative Structure–Property Relationship, *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 28–41.
[13] T. Ivanciuc and O. Ivanciuc, Quantitative Structure–Retention Relationship Study of Gas Chromatographic Retention Indices for Halogenated Compounds, *Internet Electron. J. Mol. Des*. **2002**, *1*, 94–107, http://www.biochempress.com.
[14] R. Hiob and M. Karelson, Quantitative Relationship between Rate Constants of the Gas–Phase Homolysis of N–N, O–O and N–O Bonds and Molecular Descriptors, *Internet Electron. J. Mol. Des*. **2002**, *1*, 193–202, http://www.biochempress.com.

[15] O. Ivanciuc, T. Ivanciuc, and A. T. Balaban, Quantitative Structure–Property Relationships for the Normal Boiling Temperatures of Acyclic Carbonyl Compounds, *Internet Electron. J. Mol. Des.* **2002**, *1*, 252–268, http://www.biochempress.com.
[16] HyperChem 5.1, Hypercube, Inc., Florida Science and Technology Park, 1115 N.W. 4th Street Gainesville, Florida 32601, U.S.A., E–mail info@hyper.com, www http://www.hyper.com.
[17] MOPAC 6, adapted for Windows by V. Lobanov, www http://www.ccl.net.
[18] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, AM1: A New General Purpose Quantum Mechanical Molecular Model, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[19] CODESSA 2.13, Semichem, 7204 Mullen, Shawnee, KS 66216, U.S.A., E–mail andy@semichem.com, www http://www.semichem.com.