ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/7059484

Skin Permeation Rate as a Function of Chemical Structure

ARTICLE *in* JOURNAL OF MEDICINAL CHEMISTRY · JULY 2006

Impact Factor: 5.45 · DOI: 10.1021/jm051031d · Source: PubMed

CITATIONS
27

READS
593

9 AUTHORS, INCLUDING:

Evrim Hür
Eskisehir Osmangazi University
22 PUBLICATIONS **428** CITATIONS
SEE PROFILE

Ludovic Kurunczi
Victor Babes University of Medicine and Phar…
108 PUBLICATIONS **503** CITATIONS
SEE PROFILE

Mati Karelson
University of Tartu
187 PUBLICATIONS **7,385** CITATIONS
SEE PROFILE

Alexandre Varnek
University of Strasbourg
133 PUBLICATIONS **2,203** CITATIONS
SEE PROFILE

Available from: Alexandre Varnek
Retrieved on: 29 January 2016

Skin Permeation Rate as a Function of Chemical Structure

Alan R. Katritzky,*,‡ Dimitar A. Dobchev,‡,▽ Dan C. Fara,‡ Evrim Hür,‡ Kaido Tämm,‡ Ludovic Kurunczi,§ Mati Karelson,|| Alexandre Varnek,⊥ and Vitaly P. Solov'ev⊗

Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611, Faculty of Pharmacy, University of Medicine and Pharmacy Timisoara, 2 Eftimie Murgu, Timisoara 1900, Romania, Department of Chemistry, Tallinn University of Technology, Ehitajate tee 5, Tallinn 19086, Estonia, Faculté de Chimie, 4, rue B. Pascal, Strasbourg 67000, France, Institute of Physical Chemistry, Russian Academy of Sciences, Leninskii Prosp 31, 119991, Moscow, Russia, and Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

Received October 13, 2005

Multilinear and nonlinear QSAR models were built for the skin permeation rate (Log K_p) of a set of 143 diverse compounds. Satisfactory models were obtained by three approaches applied: (i) CODESSA PRO, (ii) Neural Network modeling using large pools of theoretical molecular descriptors, and (iii) ISIDA modeling based on fragment descriptors. The predictive abilities of the models were assessed by internal and external validations. The descriptors involved in the equations are discussed from the physicochemical point of view to illuminate the factors that influence skin permeation.

1. Introduction

Transdermal delivery systems are attractive for both topical and systemic therapeutics. However, the skin barrier which protects the body from physical and chemical attacks also hinders the delivery of the required drug dose through the skin to a target organ.[1] Many exogenous substances have difficultly passing the *stratum corneum* to reach the blood circulation system. So-called skin permeation enhancers can improve the penetration of other substances by perturbing the barrier function of the *stratum corneum*.[2] Many studies have attempted to improve the permeation of drugs through the skin. The penetration of exogenous chemicals through the human skin is of significance in many disciplines, ranging from (i) the pharmaceutical and cosmetic industries, where control of permeation is essential for the topical application of lotions, creams, and ointments, to toxicological risk assessments of materials from the environment and in specific occupations.[3]

The primary barrier to percutaneous absorption is the *stratum corneum*, and the penetration of a compound is determined mainly by its chemical structure. The many additional factors which play more subtle roles in influencing percutaneous absorption include (i) the manner of application or contact of compounds to the surface of the skin, (ii) strategies which alter the barrier properties of the *stratum corneum*, (iii) processes occurring in viable tissues, and (iv) biological factors.[4]

Many efforts have been made to develop experimental approaches to measure percutaneous absorption: most available data have been obtained by in vitro diffusion chamber experiments, while the recorded in vivo data are commonly obtained via biomonitoring.[4,5] Percutaneous absorption and effects of skin penetration enhancers on the transdermal transport rate of various drugs are most frequently studied by conducting in vitro experiments, which are much simpler to implement than in vivo measurements.[6]

Permeation rates can be measured by (i) monitoring in vivo drug release in live animals or human volunteers, (ii) employing excised skin from human or animal sources, or (iii) using synthetic model membranes as diffusion barriers in in vitro experiments.[7] Human skin from various sources, including cosmetic surgery and amputations, has been used for the in vitro assessment of percutaneous penetration.[8−12] Skin from a wide range of animals including pigs, rats, guinea pigs, snakes, rabbit, and others has been suggested as a suitable replacement for human skin.[13−18] Among their species pig and rat skin barriers to diffusion are the most similar to that of human skin and hence are most commonly employed. Although animal skin provides a good indication of the diffusion characteristics of chemicals, it does not necessarily reproduce the complex nature of the human *stratum corneum*.

As a result of the problems of obtaining excised skin and those associated with implementation of synthetic membranes, there has been considerable interest in the use of quantitative structure−activity relationships (QSARs) method to predict the permeability of chemicals through the skin. This approach attempts to relate statistically the experimentally measured percutaneous penetration of a range of exogenous chemicals to known physicochemical parameters. For the development of any QSAR, a key feature is that all the "biological" data must be reproducible and consistently measured using same protocol. For skin permeation this implies skin from the same animal (and probably the same sex) and ideally measured in one laboratory and by the same workers.[19]

QSARs are useful in predicting behavior of novel compounds and providing insights into mechanisms of activity. In transdermal studies, the technique is often based on multivariate regression analysis of molecular features that determine an index of permeation such as the permeability coefficient, k_p, or the diffusion of the permeant across some part of the skin.[20] Earlier reports[21−23] based on rather small data sets revealed a linear relationship with hydrophobicity, i.e., the more hydrophobic a compound, the more rapid its skin permeation.

In 1990, Flynn[24] assembled a dataset of human skin permeability coefficients and logarithmic function of octanol−water partition coefficients, Log *P*, for over 90 compounds. He demonstrated that, for groups of molecules of a similar size,

* Corresponding author: Phone: (352) 392-0554, Fax: (352) 392-9199, e-mail:katritzky@chem.ufl.edu.
‡ University of Florida.
§ University of Medicine and Pharmacy Timisoara.
|| Tallinn University of Technology.
⊥ Faculté de Chimie.
⊗ Russian Academy of Sciences.
▽ University of Tartu.

algorithms based on Log *P* can provide a first estimate of the skin permeability coefficients. Subsequently, many QSARs for skin permeation have been published utilizing the Flynn data set, where the data have either been used as a whole or subsets taken to account for different types of compounds. The QSAR treatment of skin permeation was reviewed in detail by Moss, Dearden et al. in 2002.[25]

One aim of the present study is to relate Log K_p (the logarithm of the skin permeation rate) with theoretical descriptors (constitutional, topological, geometrical, charge related and semiempirical) calculated solely from the molecular structure. We treated a data set of 143 diverse organic compounds using both multilinear and nonlinear methodologies. The target property was related to (i) molecular descriptors, which are certain physicochemical parameters calculated either by quantum mechanical methods or by an empirical technique (CODESSA PRO software), and (ii) fragment descriptors, which use counts of different types of topological fragments of molecules as variables in a multiple regression analysis (ISIDA package [ISIDA (*I*n *S*ilico *D*esign and *D*ata *A*nalysis)],[26]). Both techniques have their advantages and disadvantages. In addition, an in-house built artificial neural network (ANN) was used for studying the nonlinear dependency between the Log K_p and theoretical molecular descriptors. The possibility of using calculated Log *P* values together with the theoretical molecular descriptors for the prediction of Log K_p was also examined.

2. Data Set

The experimental values given in Table 1 of Log K_p for the full data set of 143 compounds were collected from 10 sources.[27−36] The compilation of the data set comprised three distinct steps: (i) the conversion of (cm/hr) units into (cm/s) for K_p, (ii) the calculation of the average value, if more than one experimental value was found for a particular compound, and (iii) the estimation of the logarithm of K_p.

Since Log *P* is important for skin permeation, it was essential to develop a QSAR model for the prediction of Log *P*: to avoid the need for experimental Log *P* data for assessment of skin permeation rates. The experimental Log *P* values were taken from the literature.[37]

Table (SI-1) in the Supporting Information contains: (i) the CAS number of each compound, (ii) the experimental values of the octanol:water partition coefficient, Log *P* where available, and (ii) the calculated values of Log *P* (the last three columns).

Three software programs were assessed in the current work for the computational prediction of Log *P*:

(i) The KowWin program estimates the log octanol/water partition coefficient (Log *P*) of organic chemicals using an atom/fragment contribution method developed at Syracuse Research Corporation.[38] The calculated values of Log *P* using this method are given in the column Log P_{calc1} of SI-1.

(ii) The CLOGP program calculates Log *P* adding together values for structural parts of a solute molecule and correction factors dependent upon the particular way the parts are put together.[39] These values are given in the column Log P_{calc2} of SI-1.

(iii) The IA Log *P* and LogW Predictor,[37] developed by Interactive Analysis Inc., uses neural network implementation of E-State indices to derive a set of 10-fold cross-validated networks. The corresponding calculated values of Log *P* for all 143 organic compounds are given in the column Log P_{calc3} of SI-1.

All these three programs were accessed and used from the Interactive Analysis Inc. webpage.[37]

3. Methodology

3.1. Multilinear Regression Analysis. 3.1.1. CODESSA PRO Approach. The two-dimensional molecular structures of the 143 species were drawn using ISIS/Draw 2.4.[40] Preoptimization and three-dimensional conversions used molecular mechanics force field method (MM+) included in Hyperchem 7.0.[41] Final optimizations were performed with CODESSA PRO using the AM1 parametrization within the semiempirical quantum-chemical program CMOPAC, an in-house modification of MOPAC version 7.0.[42] The following keywords were used: AM1 VECTORS BONDS PI POLAR PRECISE ENPART EF. A gradient norm of 0.01 kcal/Å was used to test the optimized geometry for the isolated molecules. AM1 force calculations were carried out with keywords AM1 FORCE PRECISE THERMO ROT=1 to produce thermodynamic parameters. Constitutional, topological, geometrical, thermodynamic, quantum chemical, and electrostatic descriptors were calculated for the final optimized structures.[43]

The best multilinear regression (BMLR)[44] procedure was applied to obtain regression models from selected noncollinear descriptors. The BMLR selects the best two-parameter regression equation, the best three-parameter regression equation etc., based on the highest R^2 value in the stepwise selection procedure of the regression analysis. During the BMLR procedure the descriptor scales are normalized and centered automatically, and the final result is given in natural scales. This result is considered the best representation of the property in the given pool of descriptors.

A major decision in developing successive QSAR is when to stop adding descriptors to the model during the stepwise regression procedure. The lack of an adequate control in this respect leads to over-correlated equations, containing an excessive number of descriptors and difficult to interpret in terms of interaction mechanisms. A simple limitation for supervising the model expansion is the difference between the squared correlation coefficients of consecutive models (ΔR^2), as defined by eq 1, where $i = 1, 2, ..., n$, and $j = i + 1$, are the number of descriptors involved in the corresponding model.

$$\Delta R^2_{i,j} = R^2_j - R^2_i \tag{1}$$

Also, in order to monitor the possibility of chance correlation between the selected descriptors, the same eq 1 is applied to estimate ΔR^2_{cv}, where R^2_{cv} represents the cross-validated (Leave-One-Out method) correlation coefficient.

With low numbers of descriptors, this relationship has a steep ascent indicating substantial improvement of the correlation by adding descriptors. After a certain "break point", the relationship has much lower slope and the improvement of the model becomes less significant. Based on our previous experience, this occurs for $\Delta R^2 \leq 0.02$ units. The model corresponding to the break point is considered the best/optimum model.

3.1.2. ISIDA Approach. The structure−property modeling was performed using the ISIDA program which realizes the Substructural Molecular Fragments (SMF) method.[45] The SMF method is based on the splitting of a molecular graph on fragments (subgraphs) and on the calculation of their contributions to a given property *Y*. Two different classes of fragments are used: "sequences" (**I**) and "augmented atoms" (**II**). Three subtypes AB, A, and B are defined for each class. For the fragments **I**, they represent sequences of atoms and bonds (AB), of atoms only (A) or of bonds only (B). The number of atoms in these sequences varies from 2 to 6, and only the shortest paths from one atom to the other are used. An "augmented atom"

Table 1. QSPR Modeling of Skin Permeation Rate: Experimental and Predicted Log K_p Values

			predicted log K_p (cm/s)						predicted log K_p (cm/s)		
ID	compound	exp log K_p	ISIDA eq 2	CODESSA eq 7	ANN	ID	compound	exp log K_p	ISIDA eq 2	CODESSA eq 7	ANN
1	1,1,1-trichloroethane	−5.9	−5.52	−5.05	−4.53	73	ethanol	−6.53	−6.61	−6	−6.16
2	1,2,3-trihydroxybenzene	−6.37	−6.85	−6.04	−6.4	74	ethyl benzene	−3.36	−3.99	−4.25	−4.29
3	1,2,4-trihydroxybenzene[a]	−7.46	−6.85	−6.3	−6.78	75	diethyl ether [a]	−6.02	−5.52	−5.52	−5.74
4	1,3,5-trihydroxybenzene	−6.07	−6.85	−6.57	−6.91	76	ethyl nicotinate	−5.28	−5.58	−5.89	−5.74
5	12-deoxyphorbol 13-isobutyrate	−5.99	−6.18	−6.75	−7.02	77	etorphine	−5.88	−5.67	−7	−6.63
6	2,3-butanediol	−7.62	−7.71	−7.21	−7.34	78	fentanyl	−5.56	−5.85	−5.28	−4.99
7	2,4,6-trichlorophenol	−4.7	−4.49	−5.13	−5.16	79	fluocinonide	−6.33	−6.73	−6.83	−7.05
8	2,4-dichlorophenol[a]	−4.7	−4.72	−5.2	−5.59	80	hydrogencarboxylic acid	−6.6	−6.38	−5.78	−6.05
9	2-amino-4-nitrophenol	−6.67	−6.49	−6.15	−5.95	81	heptanoic acid	−5.03	−4.96	−5.26	−5.25
10	2-butanone	−5.78	−5.68	−6.03	−6.02	82	hexanoic acid	−5.18	−5.21	−5.5	−5.48
11	2-chlorophenol	−4.96	−4.95	−5.29	−5.37	83	hexyl nicotinate	−4.83	−4.55	−4.82	−4.74
12	2-ethoxyethanol	−7.06	−6.61	−6.51	−6.68	84	hydrocortisone hemipimelate	−6.3	−6.27	−6.44	−6.6
13	2-heptanol	−5.04	−5.33	−5.23	−5.20	85	hydrocortisone hemisuccinate	−6.76	−7.04	−7.32	−7.67
14	2-hydroxypropyl nicotinate	−7.55	−6.21	−6.11	−6.37	86	hydrocortisone hexanoate	−5.3	−5.80	−5.82	−5.63
15	2-nitrophenol	−4.08	−5.38	−5.44	−5.44	87	hydrocortisone octanoate	−4.77	−5.29	−5.23	−5.02
16	2-nitro-*p*-phenylenediamine	−6.81	−6.78	−6.78	−6.36	88	hydrocortisone propinate[a]	−6.17	−6.57	−6.83	−6.47
17	2-phenylethanol[a]	−5.23	−5.08	−5.63	−6.04	89	hydrocortisone	−7.85	−8.06	−7.76	−7.84
18	3,4-xylenol[a]	−4.92	−4.84	−5.1	−5.54	90	hydromorphone	−8.03	−8.24	−7.31	−7.49
19	3-phenylpropanol	−4.84	−4.83	−5.45	−5.48	91	hydroquinone[a]	−6.51	−6.01	−6.01	−6.52
20	4-amino-2-nitrophenol	−6.04	−6.49	−6.37	−6.23	92	hydroxypregnenolone[a]	−6.78	−6.39	−6.21	−5.63
21	4-bromophenol	−4.92	−5.38	−4.99	−5.04	93	hydroxyprogesterone[a]	−6.78	−6.18	−6.77	−6.09
22	4-chloro-3-methylphenol[a]	−4.71	−4.78	−4.89	−5.31	94	ibuprofen[a]	−5.3	−5.16	−4.88	−5.28
23	1-chloro-2,4-diaminobenzene[a]	−6.31	−6.35	−6.73	−6.58	95	indomethacin	−5.85	−5.85	−5.1	−5.42
24	4-ethylphenol[a]	−4.93	−4.82	−4.87	−5.29	96	isoquinoline	−5.29	−5.07	−5.31	−5.03
25	4-hydroxybenzyl alcohol	−6.26	−6.47	−6.42	−6.71	97	lidocaine	−6.17	−6.17	−5.82	−5.59
26	4-hydroxy-methylphenylacetate[a]	−5.26	−5.75	−5.73	−6.29	98	1-hydroxy-3-methylbenzene	−5.29	−5.01	−5.09	−5.19
27	4-hydroxyphenylacetamide	−6.89	−7.38	−6.89	−6.95	99	meperidine	−5.7	−5.66	−5.61	−5.56
28	4-hydroxyphenylacetic acid	−6.06	−6.23	−6.38	−6.65	100	methanol	−6.7	−6.99	−6.03	−6.4
29	4-phenylbutanol	−4.74	−4.57	−5.28	−5.29	101	methyl 4-hydroxybenzoate	−5.44	−5.28	−5.31	−5.53
30	5-fluorouracil	−6.82	−6.76	−7.91	−7.3	102	methyl nicotinate	−5.77	−5.95	−6.24	−6.11
31	5-phenylpentanol[a]	−4.53	−4.31	−4.92	−5.3	103	3-hydroxy-1-nitrobenzene	−5.73	−5.38	−5.34	−5.35
32	6-phenylhexanol	−4.56	−4.06	−4.7	−4.73	104	morphine[a]	−8.3	−8.07	−7.64	−7.6
33	8-methoxypsoralen	−5.12	−4.95	−5.63	−6.01	105	naproxen	−5.74	−6.31	−5.15	−5.41
34	acetic acid	−6.53	−6.94	−6.11	−6.19	106	1-butanol	−6.08	−6.10	−5.75	−5.79
35	aldosterone	−8.17	−8.25	−8.05	−8.13	107	1-decanol	−4.47	−4.56	−4.02	−4.27
36	amylobarbital	−6.14	−6.14	−6.29	−6.35	108	1-heptanol	−4.91	−5.33	−5.02	−5.01
37	aniline[a]	−5.05	−5.46	−5.69	−5.81	109	1-hexanol[a]	−5.27	−5.59	−5.29	−5.59
38	anisole	−4.8	−4.92	−4.84	−4.86	110	nicotine[a]	−5.47	−5.47	−6.53	−6.65
39	aspirin	−5.69	−5.69	−5.97	−6.19	111	nitroglycerine	−5.44	−5.52	−5.83	−5.62
40	atropine	−7.8	−7.36	−6.59	−6.91	112	1-octanol	−4.69	−5.07	−4.88	−4.87
41	barbital	−7.47	−7.40	−7.18	−7.22	113	1-nonanol[a]	−4.68	−4.82	−4.47	−4.88
42	benzaldehyde[a]	−4.47	−4.54	−5.28	−5.69	114	1-pentanol	−5.7	−5.84	−5.5	−5.52
43	benzene[a]	−4.41	−4.35	−4.36	−4.78	115	1-propanol	−6.32	−6.36	−5.94	−6.03
44	benzoic acid	−5.15	−4.93	−5.24	−5.35	116	1-hydroxy-2-methylbenzene	−5.28	−5.01	−5.08	−5.14
45	benzyl alcohol	−5.41	−5.64	−5.58	−5.64	117	1-heptanecarboxylic acid[a]	−4.93	−4.70	−4.92	−5.27
46	benzyl nicotinate[a]	−4.87	−4.60	−5.45	−5.72	118	1,2-benzenediamine	−6.85	−6.58	−6.68	−6.32
47	naphthalene-2-ol	−5.04	−4.40	−4.84	−5.32	119	ouabain	−9.66	−9.79	−10.11	−9.26
48	bromodichloromethane	−3.82	−3.82	−4.14	−4.2	120	4-hydroxychlorobenzene[a]	−4.92	−4.95	−5.08	−5.52
49	bromoform	−3.75	−3.75	−3.58	−3.9	121	1-hydroxy-4-methylbenzene[a]	−5.23	−5.01	−5.09	−5.51
50	butobarbital	−7.23	−6.89	−6.56	−6.6	122	pentanoic acid	−5.96	−5.47	−5.71	−5.7
51	butyl nicotinate	−4.86	−5.06	−5.47	−5.31	123	phenobarbital[a]	−6.66	−6.68	−6.69	−6.96
52	butyric acid[a]	−6.27	−5.73	−6	−6.2	124	phenol	−5.52	−5.18	−5.14	−5.28
53	caffeine	−6.46	−6.46	−7.6	−7.36	125	piroxicam	−7.37	−7.43	−6.04	−6.14
54	catechol	−6.07	−6.01	−5.85	−6.12	126	4-hydroxynitrobenzene	−5.73	−5.38	−5.36	−5.38
55	chlorodibromomethane	−3.78	−3.78	−3.89	−4.05	127	1,4-benzenediamine	−7.13	−6.58	−6.97	−6.65
56	trichloromethane[a]	−3.92	−3.92	−4.36	−4.63	128	pregnenolone[a]	−6.38	−6.39	−6.04	−5.29
57	chloroxylenol[a]	−4.72	−4.61	−4.95	−5.35	129	progesterone[a]	−5.5	−5.79	−6.31	−5.54
58	chlorpheniramine	−6.19	−6.21	−5.69	−5.27	130	propionic acid[a]	−6.02	−5.98	−6.06	−6.25
59	codeine	−7.58	−7.81	−7.35	−7.45	131	resorcinol	−7.03	−6.01	−5.93	−6.24
60	cortexolone	−7.68	−6.71	−6.84	−6.74	132	salicylic acid	−5.53	−5.77	−5.2	−5.47
61	cortexone	−6.9	−6.32	−6.99	−6.8	133	scopolamine	−7.87	−7.88	−7.19	−7.68
62	corticosterone	−7.26	−7.67	−7.56	−7.56	134	styrene	−3.75	−4.54	−4.19	−4.27
63	cortisone[a]	−8.56	−7.84	−7.89	−7.5	135	sucrose	−8.84	−8.84	−9.29	−9.19
64	dexamethasone[a]	−7.53	−7.36	−7.72	−7.28	136	sufentanyl	−5.4	−5.40	−5.66	−5.27
65	diclofenac	−5.73	−6.00	−4.96	−4.94	137	testosterone	−6.27	−6.32	−6.51	−6.18
66	diethyl ether[a]	−5.37	−5.52	−5.52	−5.74	138	tetrachloroethylene	−4.86	−5.52	−4.34	−4.2
67	diethylcarbamazine	−7.01	−7.15	−7.2	−7.1	139	thymol	−4.74	−5.21	−4.66	−4.7
68	digitoxin	−8.37	−8.14	−7.77	−7.61	140	toluene	−3.58	−4.18	−4.28	−4.33
69	ephedrine	−5.72	−5.64	−6.29	−6.19	141	trichloroethylene	−4.03	−4.03	−4.62	−4.61
70	estradiol[a]	−6.14	−6.59	−5.64	−5.48	142	urea[a]	−7.39	−7.14	−7.91	−7.24
71	estriol[a]	−7.96	−7.94	−6.73	−6.77	143	water	−6.53	−6.53	−6.36	−6.77
72	estrone	−6	−5.93	−6.23	−6.22						

[a] These 40 compounds were randomly selected as the validation set for the ANN approach and were not included into the ANN training set.

represents a selected atom with its environment including both neighboring atoms and bonds (AB), or atoms only (A), or bonds only (B). Atomic hybridization (Hy) can be taken into account for augmented atoms of the A-type. Benson type augmented

atoms **II**(Benson) correspond to substructural fragments suggested in ref 45 for fast estimation of enthalpy of formation and the enthalpy of organic compounds within an additive scheme.

Once a molecular graph is split into constitutive fragments, any corresponding quantitative physical or chemical property Y is calculated from the fragment contributions using linear (eq 2) or nonlinear (eq 3) and (eq 4) fitting equations.

$$Y = a_o + \sum_i a_i N_i + \Gamma \quad (2)$$

$$Y = a_o + \sum_i a_i N_i + \sum_i b_i (2N_i^2 - 1) + \Gamma \quad (3)$$

$$Y = a_o + \sum_i a_i N_i + \sum_{i,k} b_{ik} N_i N_k + \Gamma \quad (4)$$

where, a_i and b_i (b_{ik}) are fragment contributions, N_i is the number of fragments of i type. The a_o term is fragment independent. The a_o, a_i, and b_i (b_{ik}) are obtained by the multilinear regression procedure using the training set of compounds. An extra term $\Gamma = \Sigma c_m D_m$ can be used to describe any specific feature of the compound using external descriptors D_m (e.g., topological, electronic, etc.), by default $\Gamma = 0$.

At the training stage, builds up to 147 structure−property models involving 49 types of fragment descriptors and the three linear and nonlinear fitting equations. If some fragments are linearly dependent, they are treated as one extended fragment. Using the singular value decomposition method (SVD) the program fits the a_i and b_i terms in eqs 2−4 and calculates the corresponding statistical characteristics (correlation coefficient (R), standard deviation (s), Fischer's criterion (F), cross-validation correlation coefficient (Q), standard deviation of predictions (s_{PRESS}), Kubinyi's criterion (FIT), R_H-factor of Hamilton, and matrix of pair correlations (covariation matrix for the terms a_i and b_i) and performs statistical tests[24] to select the best models.

If some of the variables in eqs 2−4 are linearly dependent, or if a given fragment occurs in a relatively small number of molecules, the standard deviation Δa_i (Δb_i) for the fragment contributions a_i (b_i) can be large and lead to the corresponding t-test ($t = a_i/\Delta a_i$) being smaller than the tabulated value (t_0). In such cases the following procedure is applied in order to improve the robustness of the models. First, it selects the variable with the smallest $t < t_0$, and then it performs a new fitting excluding that variable. This procedure is repeated until $t \geq t_0$ for all variables.

ISIDA recognizes nine different types of bonds: single, double, triple (each of these in a ring or in a chain), aromatic bonds, and two types of coordination bonds. Therefore, the *EdChemS* editor of 2D structures incorporated in the ISIDA package is used to normalize the bond types in the aromatic fragments originally presented as Kekule structures and to modify bond types in rings accounting for the difference between chains and rings. Then, using the *SDF Editor* also included in ISIDA, structure data files (SDF) containing all 143 structures names of the compounds and the experimental values of Log K_p or have been prepared and further used in structure−property modeling. All hydrogen atoms were omitted in ISIDA calculations.

3.2. Nonlinear Analysis. Artificial Neural Network (ANN). An artificial neural network is a biologically inspired computer algorithm designed to treat the data in a manner emulating the learning pattern in the brain. The computer-based network accepts a set of input values, transforms these, and generates an associate set of output values.[46] Through an iterative "learning" process, the network refines the information derived from the input values (descriptors) in order to reproduce an associated set of target values (experimental Log K_p). Once a network has been trained to recognize the underlying theme for a given set of input/target pairs, it may be used to predict an output value corresponding to a new group of input values.

In our case, the signal emerging from the output neuron represents the current log skin permeation rate for a given p-th input, o_{pk}, for the set of input descriptors utilized and given the architecture with k outputs of the network. When this value is compared to the desired (experimental target) Log K_p, t_{pk}, a measure of the network error can be calculated. The sum-squared error used to quantify the effectiveness of the network is given by eq 5. The data are repeatedly passed through the network, with the overall error successively decreasing as the network adjusts the weights and biases to reflect the structure−skin permeation relationship.

$$E = \sum_1^p (t_{pk} - o_{pk})^2 \quad (5)$$

The back-propagation network and the delta rule for optimization of the weights were used.[46] On the basis of the steepest descent method, this technique optimizes the connection weights by proceeding in the direction that most reduces the error of the estimate. The connection weight, w_{ij}, is corrected by the amount Δw_{ij}, which is proportional to the network error E with respect to each weight as shown in eq 6, where η is the learning rate set by the researcher. The learning rate is introduced in order to control the average step size of the weight change. In the training process the weights are iteratively adjusted by appropriate amount (eq 6).

$$\Delta w_{ij} = -\eta \frac{\partial E_p}{\partial w_{jk}} \quad (6)$$

A neural network is trained, i.e. the neural network learns from a predefined group of compounds known as the training set. One potential problem associated with the use of network technique is that it may overtrain i.e. derive a relationship which is too specialized. This problem is averted by delegating a portion of the compounds to serve as "a validation set". Because this set of compounds has no direct influence on the actual learning process, it can be used to monitor the predictive capability of the network at regular intervals during a training run. By tracking the validation set error, the optimal set of weights (and biases) to be used as the final predictive model can then be identified.

3.3. Internal Validation. To perform internal validation of the multilinear QSPR (CODESSA PRO) models and the models obtained by ISIDA, the parent data set was divided into three subsets (a, b, c): the 1st, 4th, 7th, etc. data points go into the first subset (a), the 2nd, 5th, 8th, etc. into the second subset (b), and the 3rd, 6th, 9th, etc. into the third subset (c). Then, three training sets A, B, and C were prepared as the combinations of two subsets (a and b), (a and c), and (b and c), respectively. The remaining subsets (c, b, and a, respectively) become the corresponding test sets.

For each of the training sets the correlation equation was derived with the same descriptors in eq 7. Then, the obtained equation was used to predict Log K_p values for the compounds from the corresponding test set.

Table 2. Selection of the Best Sets of Descriptors. Best Models

	POOL 1		POOL 2		POOL 3	
	R^2	ΔR^2	R^2	ΔR^2	R^2	ΔR^2
2	0.7		0.7			
3	0.75	0.04	0.75	0.04	0.76	
4	0.77	0.03	0.77	0.03	**0.81**	**0.05**
5	**0.81**	**0.03**	**0.81**	**0.03**	0.84	0.04

The efficiency of QSAR models to predict Log K_p value was estimated using the cross-validation (Leave One Out method[47]) correlation both for the full set and for each of the three training sets. The correlation coefficients and standard deviations of linear correlations between experimental and predicted for test sets Log K_p values were also calculated.

4. Results

4.1. CODESSA PRO Calculations. Our attempt to build a useful QSAR tool for prediction of skin permeation rate used various combinations of the existing descriptors for different types of correlations.

(1) POOL 1 (878 parameters) (CODESSA): Models based only on molecular descriptors (877) and Log *P* as external descriptor;

(2) POOL 2 (1181 parameters) (CODESSA + ISIDA): Models with molecular descriptors (877), fragment descriptors (303), and Log *P*;

(3) POOL 3 (1183 parameters) (ANN): Models with molecular descriptors (877), fragment descriptors (303), and Log *P*, $(\log P)^2$, $(\text{Log } P)^{0.5}$.

To find the model with the optimum set of descriptors, multiple parameter correlations involving up to 10 descriptors were developed for each of the three descriptors sets used. The "break point" algorithm described in section 3.1.1 was then applied to determine the best correlation (see Table 2). In addition, to reduce the possibility of chance correlations for the descriptors selected in the equations, the cross-validated correlation coefficient R^2_{cv} was monitored for significant increases on increasing the number of descriptors (see eq 1). Also, a randomization Monte Carlo test[50] for chance correlations for each equation indicated that the possibility of chance correlation was less than 0.001.

The algorithm for descriptor selection for the ANN model is different than the above-mentioned and is described in section 4.3. However, for completeness we also present the results for ANN in Table 2.

First, the information given in Table 2 reveals that the best sets of descriptors have five parameters, when the 1st or 2nd pool of descriptors was used for multilinear regression, and four parameters, when the 3rd pool (neural network approach) was used. Comparison of the five-parameter models developed for POOL 1 and POOL 2 shows that they are identical. Importantly, all pools include Log P_{exp} as external descriptor, and Log P_{exp} is also present in the five-parameter CODESSA model. Obviously, Log P_{exp} relates significantly to the skin permeation rate. The four remaining parameters are all theoretical descriptors calculated solely from the molecular structure.

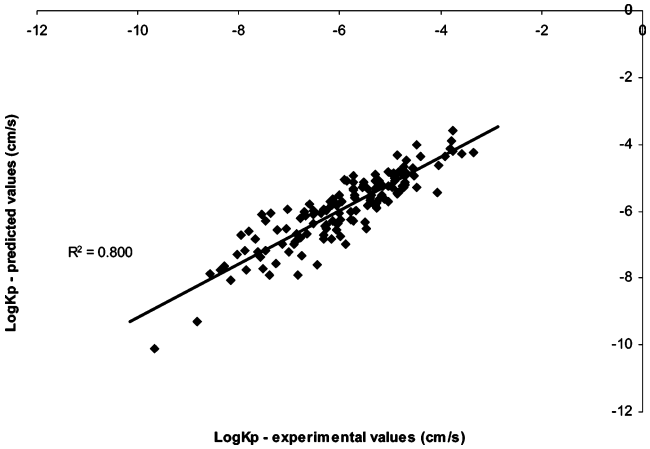

Figure 1. Predicted vs experimental Log K_p values by using eq 7, CODESSA PRO approach (with Log P_{exp}).

On the basis of the above-mentioned observations, we selected the five-parameter regression equation as best QSAR model developed using the CODESSA PRO approach. The mathematical form of this QSAR model, given by eq 7, has the following statistical characteristics: $N = 143$, $k = 5$, $R^2 = 0.800$, $R^2_{cv} = 0.781$, $F = 109.6$, and $s = 0.54$.

$$\text{Log } K_p = -(7.20 \pm 0.17) + (0.67 \pm 0.04)D_1 - (0.24 \pm 0.03)D_2 + (0.74 \pm 0.09)D_3 - (23.43 \pm 3.50)D_4 + (0.20 \pm 0.04)D_5 \quad (7)$$

The symbolic notations of descriptors involved in the QSAR models are presented in Table 3.

The predicted and experimental Log K_p values are given in Table 1. The correlation between predicted and observed values for the Log K_p (eq 7) is plotted in Figure 1.

4.2. ISIDA Calculations. Calculations on the full set using the ISIDA approach resulted in a statistically significant linear model based on the **II**(Benson) fragments which represent augmented atoms built according to the scheme suggested in

Table 3. Molecular Descriptors Involved in the QSAR Models

			models		
			CODESSA PRO		
	descriptor name	symbol	eq 7	eq 13	ANN
		External			
1	Log P_{exp}	D_1	D_1		D_1
2	Log P_{calc1}	D_7		D_7	
		Topological			
3	Kier & Hall index (order 3)	D_2	D_2	D_2	D_2
		Thermodynamic			
4	rotational entropy (300 K)/no. of atoms	D_3	D_3	D_3	D_3
		Charge-Distribution-Related			
5	HASA-2/TMSA (Zefirov Partial Charge)	D_4	D_4	D_4	
6	H-acceptors FCPSA (version 2)	D_6			D_6
		Constitutional			
7	number of O atoms	D_5	D_5	D_5	

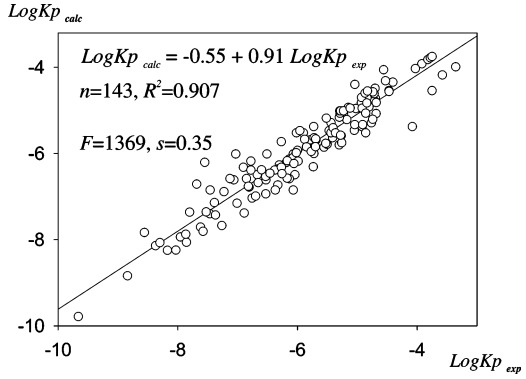

Figure 2. Calculated vs experimental skin permeation rate (Log K_p). ISIDA calculations were performed for the full sets of compounds using linear **II**(Benson) model.

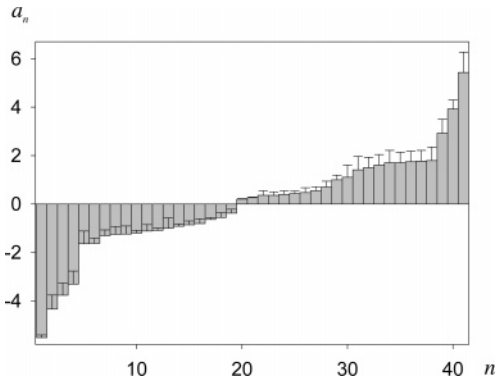

Figure 3. Modeling of skin permeation rate by ISIDA program. Contributions (a_n) of the fragments to skin permeation rate (Log K_p) for the **II**(Benson)/eq 1 model for the full data set. The standard deviations are given as error bars. The fragments numbers (*n*) correspond to the following fragments listed in Table 4.

ref 45. The statistical parameters of this model ($N = 143$, $k = 41$, $R^2 = 0.907$, $s = 0.43$, $R_{cv}^2 = 0.812$, $F = 24.8$) are considerably better than those obtained for the CODESSA PRO models. The plot of calculated vs experimental Log K_p values (Figure 2) also demonstrates good performance of the model. The fragments' contributions are given in Figure 3.

4.3. ANN Calculations. Before the neural network treatment was started, the experimental property values and descriptor values were both normalized to a range 0−0.9. The significant descriptors were selected from the large descriptor pool. The available data were divided into a training and a validation set. POOL 3 was used for the descriptor selection. The validation set for the network was constructed from almost one-third (40) of the compounds randomly selected from the full set. This provides a check against over-training of the models (these compounds are highlighted in Table 1). The remaining (103) compounds were used as the training set for the network. All CODESSA PRO descriptors were examined for intercorrelations and redundant and highly correlated descriptors excluded. Further, by constructing simple 1-1-1 neural networks, sensitivity analyses were performed, and then the descriptors that produced the lowest error (5) were selected. Several neural network models with different architecture were investigated using the descriptors selected as above. The best model that was found was 4-4-1, i.e. four neurons (descriptors) in the input layer, four neurons in the hidden layer, and one neuron in the output layer. The input neurons consist of the following descriptors*:* Kier & Hall index (order 3), Rotational entropy (300K), H-acceptor FCPSA (version 2), and Log P_{exp}. The latter was forced into the model since it showed a good correlation with the property under investigation. The next stage of the computation was to train the neural network. By starting with the random weights, the optimized weights were obtained after (approximately) 8000 epochs (cycles). For each epoch, the weights were updated with the amount defined by eq 6.

During the training (optimization), the output error (RMS) was tracked using R^2 of the validation set. When the output error started increasing (after a certain epoch), the training process was stopped and the optimized weights were selected as final. The optimal weights were obtained at maximum $R^2 = 0.721$ and minimum RMS (root-mean squared error) = 0.661 of the validation set. At this point the RMS minimum was achieved for the validation set, and the ANN training was stopped to prevent over-fitting. The results for the predicted Log K_p values of the training set are shown in Table 1. The predicted vs experimental Log K_p for the training set are shown as a linear fit in Figure 4.

The RMS and R^2 values for the training set were 0.519 and 0.813, respectively. Similarly, the predicted vs experimental Log K_p for the validation set are shown in Figure 5. The validation set was not used for training the network; thus, it serves also as external prediction for this neural network model.

The comparison of the linear and ANN models indicates that the ANN models are slightly better than the linear models (see Table 2). However, the models are built on different number of compounds and therefore the direct comparison is not straightforward. In our neural network modeling, it was implied that the model has to be as simple as possible and to have as good prediction as possible. With only four descriptors in the input, the network was able to predict the data for 40 additional compounds with a reasonable correlation of $R^2 = 0.72$. In addition, the standard deviation of the ANN model ($s = 0.524$)

Table 4. Fragment Numbers and the Corresponding Fragments

n	fragment	*n*	fragment	*n*	fragment	*n*	fragment
1	a_0 [a]	11	C(C′C$_B$′CO)	21	C(C′C)	31	C$_B$(S$_B$) [f]
2	C(C′C′C′CO) [b]	12	O(C)	22	C$_D$(CO)	32	C$_D$(Cl)
3	C(C′C′C′C)	13	O	23	C(C$_B$) [e]	33	C(Cl′Cl′Cl)
4	C(C′C′C′C$_B$)	14	C$_B$(N)	24	C(C′CO)	34	C(C′N′C$_B$)
5	CO(C$_B$′N$_I$)	15	O(CO)	25	C$_B$(Cl)	35	C(Cl′Cl′Br)
6	N(C′C′C)	16	N(CO)	26	CO(O′C$_B$)	36	C(Cl′Br′Br)
7	N$_I$(C$_B$) [c]	17	O(C$_B$)	27	C(C′C$_B$)	37	C(Br′Br′Br)
8	CO(C′C$_D$)	18	C(CO)	28	C(C′C′O′CO)	38	C(N′CO)
9	CO(N′C$_D$) [d]	19	C(O)	29	C(C′C′C)	39	C(C′O′C$_D$)
10	CO(C′N)	20	C$_B$	30	CO(N′C$_B$)	40	C(C′C′C′O)
						41	C(C′C′O′O)

[a] The a_o term is fragment independent. [b] C(C′C′C′CO) is a C_{sp3} atom connected to three C_{sp3} atoms and one C═O group. [c] N$_I$(C$_B$) is a N_{sp2} (imino or aromatic nitrogen) atom connected to aromatic carbon. [d] CO(N′C$_D$) is a carbonyl group C═O connected to N_{sp3} and carbon atom in alkenyl group. [e] C(C$_B$) is a C_{sp3} atom connected to one aromatic carbon and three hydrogen atoms. [f] C$_B$(S$_B$) is an aromatic carbon connected to sulfur S(II) in the thiophene ring.

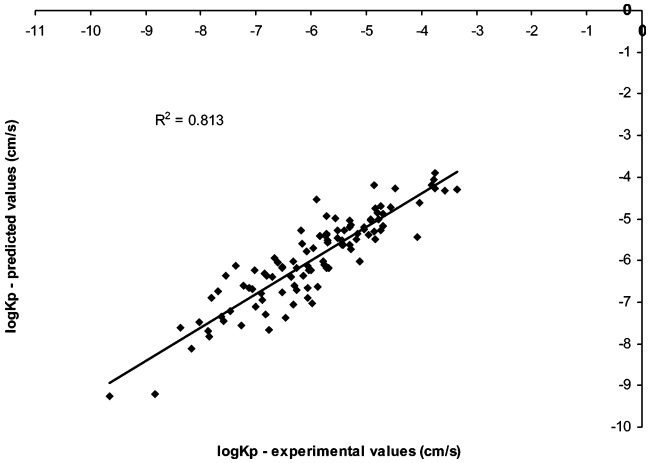

Figure 4. Predicted vs experimental Log K_p values for the training set (103 datapoints), ANN approach.

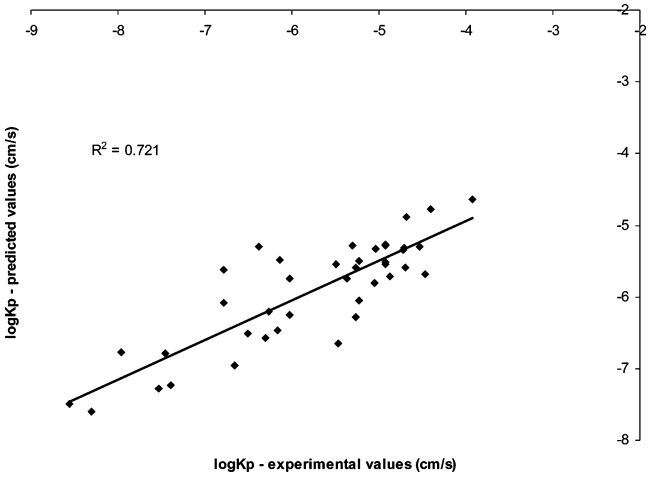

Figure 5. Predicted vs experimental Log K_p values for the test set (40 data points), ANN approach.

is less than the standard deviation of the multilinear eq 7 (s = 0.540). Also, the advantage of the ANN model is that it was already validated externally.

4.4. Validation of the Models. The validation calculations indicate: (i) The predicted R^2 values for CODESSA PRO models are in good agreement with the original QSAR model (eq 7) with the average correlation coefficients of 0.804 and 0.783 for the training and prediction sets, respectively. (ii) The selected II (Benson)/eq 1 model predicts reasonably well the modeled property for all three test sets: the average value of R^2(pred) = 0.630 is not very high, but still acceptable (see Table 5).

5. Discussion

This section contains three main topics: (i) a description and interpretation of the molecular descriptors involved in CODESSA PRO (eq 7) and ANN models, (ii) an alternative QSAR model for Log K_p that involves Log $P_{calculated}$ instead of Log $P_{experimental}$, and (iii) a comparison of the various models proposed.

As described in Table 3, the QSAR model defined by the regression eq 7 has four CODESSA PRO molecular descriptors (D_2, D_3, D_4, and D_5), and one "external" descriptor, D_1, as the logarithmic function of the experimentally measured values of the octanol:water partition coefficient, Log P_{exp}.

The most important descriptor in eq 7 according to "t-test" values is Log P_{exp} ($|t|$ = 43.12). Overall, the significance of the descriptors involved in the eq 7 model decreases in the following order: $D_1 > D_2 > D_3 > D_4 > D_5$.

Table 5. Modeling of the Skin Permeation Rate (Log K_p): Validation of the CODESSA PRO and ISIDA Models[a]

training set	n	k	R^2	s	test set	k_{test}	R^2 (pred)	s (pred)
CODESSA PRO (eq 7)								
a+b	96	5	0.802	0.55	C	5	0.786	0.47
a+c	95	5	0.808	0.48	B	5	0.779	0.59
b+c	95	5	0.802	0.58	A	5	0.784	0.45
average			0.804	0.54			0.783	0.50
ISIDA (eq 2)[b]								
a+b	96	32	0.915	0.42	C	35	0.675	0.6
a+c	95	28	0.921	0.36	B	34	0.612	1.02
b+c	95	35	0.920	0.45	A	37	0.604	1.28
average			0.919	0.41			0.630	0.97

[a] See text about the preparation of training and test sets. Statistical parameters for the sets: the number of compounds in training set (n), correlation coefficient (R) and standard deviation (s) for the model, the number of fitted coefficients in test set (k_{test}), correlation coefficient R (pred) and standard deviation s (pred) for the linear correlation Log(K_p pred) = a × Log(K_p exp) + b between experimental and predicted skin permeation rate. [b] For some compounds (from 10 to 13) from the test set, the skin permeation rate was not calculated if compounds contain unique molecular fragment(s), which were not presented in the corresponding training set.

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_i^v - 1} \tag{8}$$

The Kier & Hall index (order 3), D_2, is a valence connectivity index and accounts for the presence of heteroatoms and the hybridization of atoms in the molecule. Theoretically, D_2 is based on the atomic valence connectivity for the ith atom in the molecular structure, which is given by the mathematical formula of eq 8, where Z_i is the total number of electrons in the ith atom, Z_i^v is the number of valence electrons, and H_i is the number of hydrogen atoms directly attached to the ith nonhydrogen atom. The "order 3" classification denotes that there are three contiguous bond fragments and higher order paths, clusters, path-clusters, or chain fragments. Consequently, the definition of this descriptor is given by eq 9, where m = 3, and the summation is carried out over all subgraphs of order 3 in the entire molecule.

$$D_2 = \sum_{i=1}^{N_s} \prod_{k=1}^{m+1} \left(\frac{1}{\delta_k^v} \right)^{1/2} \tag{9}$$

The third descriptor in the model (eq 7), D_3, is the rotational entropy of the molecule at 300 K (S_{rot}) divided by the number of atoms (NA) in it, eq 10, where the rotational entropy at T = 300 K is given by eq 11.

$$D_3 = \frac{S_{rot}}{NA} \tag{10}$$

$$S_{rot} = Nk \ln \left[\frac{\pi^{1/2}}{\sigma} \prod_{j=1}^{3} \left(\frac{8\pi^2 I_j kT}{h^2} \right)^{1/2} \right] \tag{11}$$

The terms in eq 11 are: I_j, the principal moments of inertia of the molecule; σ, the symmetry number of the molecule; h, the Planck constant, and kT, the Boltzmann temperature. Both D_2 and D_3 descriptors may be related to the size and shape of the molecule.

The only charge-related descriptor in eq 7 is D_4. This parameter defines the area-weighted surface charge of the

hydrogen-bonding acceptor atoms in the molecule, HASA2, normalized by the total molecular surface area, TMSA. The mathematical expression of HASA2 is given by eq 12.

$$\mathrm{HASA2} = \sum_{A} \frac{q_A \sqrt{S_A}}{\sqrt{S_{tot}}} \qquad (12)$$

In eq 12, q_A is the partial charge on the hydrogen-bonding acceptor (H) atom(s), S_A is the surface area for this atom, and S_{tot} is the total molecular surface area. The electronegative atoms with lone electron pairs are usually considered as potential hydrogen-bonding acceptors in the molecule (O, N, etc.). As can be seen from eq 7, this descriptor has a very strong negative influence over the skin permeation rate.

The final descriptor in the five-parameter QSAR model eq 7 is D_5 that counts for the *number of O atoms in* the molecule. It is usually related to the hydrogen-bonding ability of the molecules. However, the positive sign of the regression coefficient indicates that the respective term may represent a relatively small correction to the large negative hydrogen-bonding term involving D_4. It has been noticed, that such corrections to charge-related descriptors connected to the number of certain atoms in the molecule may be due to the deficiency of standard semiempirical quantum-chemical parametrizations to describe accurately the charge distribution of molecules in condensed media. The adjustment of the parametrization for liquid media can remove such "correction terms".[48]

It was observed that descriptors D_2 and D_4 decrease the logarithm of the skin permeation rate, whereas descriptors D_1, D_3, and D_5 make positive contributions to the studied property.

The fragment contributions (Table 4) estimated by ISIDA bring useful information concerning the influence of particular molecular fragments on the skin permeation rate. Thus the quaternary carbon atom having as neighbors either three carbons and one oxygen atom (C(C′C′C′O) or two carbons (C_{sp3} and C_{sp2}) and one oxygen atom C(C′O′C_D), and the carbon atom C(N′CO) connected to one nitrogen, to one carbonyl group and to polyhalogenated carbon atoms, significantly contribute to Log K_p (Figure 3). The information about fragment contributions could be used for the preparation of focused virtual combinatorial libraries and its screening.[49,50]

The neural network model for the Log K_p consisted of the following descriptors: Kier & Hall index (order 3), Rotational entropy (300K), H-acceptor FCPSA (version 2), and Log P_{exp}. Importantly, two of these descriptors (D_1, Log P; and D_2, Kier & Hall index (order 3)) in the neural network model are also included in the best five-parameter linear model (see eq 7 and Table 3). The other two descriptors in the ANN model are very similar to two of the descriptors in eq 7: (i) Rotational entropy (300 K)/no. of atoms and Rotational entropy (300 K) (D_3), (ii) H-acceptor FCPSA (version 2) and HASA-2/TMSA (Zefirov Partial Charge) (D_4), respectively. This similarity of the descriptors in the two models is not surprising. The linear model brings up the descriptors that are the "best" in a given descriptor space. The ANN treatment accounts additionally for the nonlinearity of the dependence between Log K_p and the descriptors and thus gives slightly better description of the experimental data. Therefore, the physisco-chemical picture based on the linear model should be identical to the interpretations of the ANN model.

Further, the five-parameter QSAR models that contain the same set of four molecular descriptors (D_2, D_3, D_4, and D_5) and the calculated values of Log P as the 5th parameter (D_7), instead of Log P_{exp}, were examined.

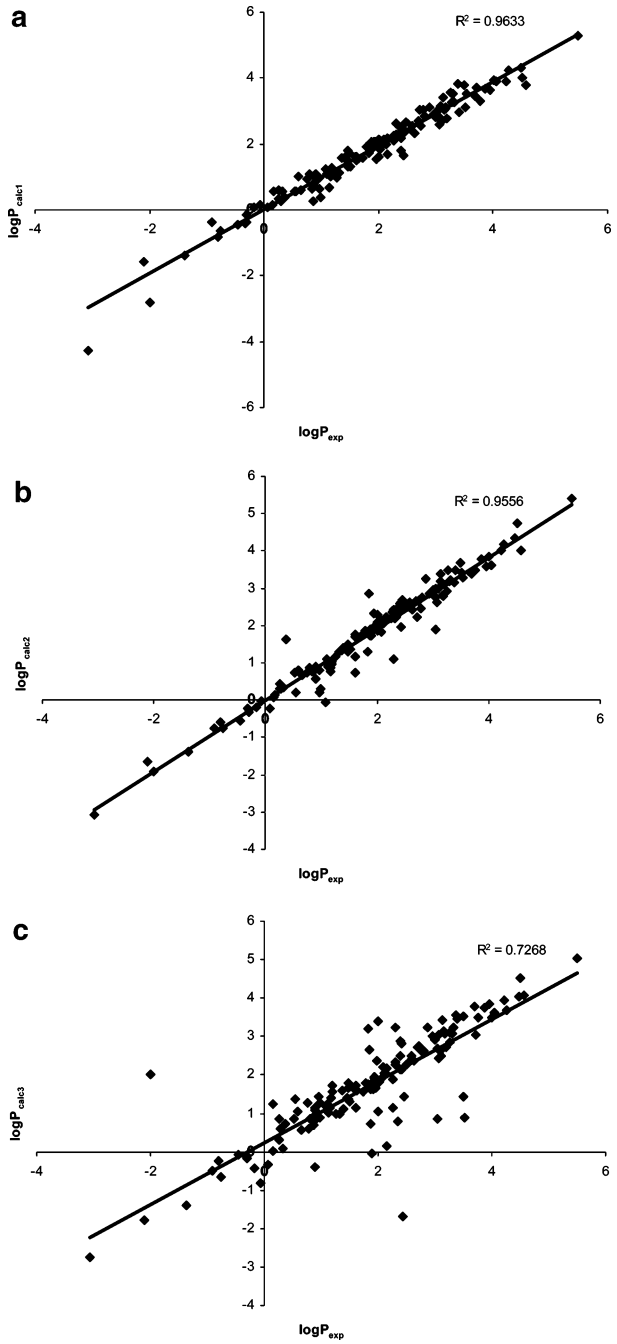

Figure 6. a. Calculated Log P_{calc1} values (ref 38) vs experimental Log P_{exp}. b. Calculated Log P_{calc2} (ref 37) values vs experimental Log P_{exp}. Figure 6c. Calculated Log P_{calc3} (ref 39) values vs experimental Log P_{exp}.

Estimation and Utilization of Log P_{calc}. Three different methods were examined for computing Log P_{calc} values (see SI-1 and Figure 6). The best correlation between the Log P_{calc} and experimental Log P values was obtained when the values of Log P were calculated using KowWin software[38] (see Figure 6a).

We therefore substituted Log P_{exp} with Log P_{calc1} in the pool of descriptors and rebuilt the QSAR model to obtain eq 13.

$$\begin{aligned}\mathrm{Log}\ K_p = &-(7.233 \pm 0.180) + (0.662 \pm 0.042)D_7 - \\ &(0.237 \pm 0.027)D_2 + (0.732 \pm 0.097)D_3 - \\ &(21.592 \pm 3.711)D_4 + (0.200 \pm 0.044)D_5 \quad (13)\end{aligned}$$

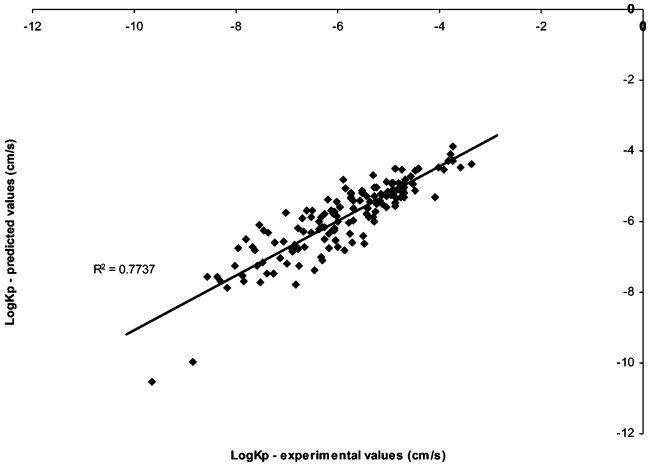

Figure 7. Predicted vs experimental Log K_p values by using eq 13, CODESSA PRO approach (with Log P_{calc1}).

Equation 13 has the following statistical characteristics: $N = 143$, $k = 5$, $R^2 = 0.774$, $R^2_{cv} = 0.747$, $F = 93.7$, and $s = 0.575$.

The correlation between predicted and observed values for the Log K_p (eq 13) is plotted in Figure 7. The main observation is that both QSAR models (eq 7 and eq 13) have similar statistical characteristics. Consequently, we assume that eq 13 can be successfully used to predict the rate of skin permeation for compounds with no available experimental values of the octanol:water partition coefficient.

6. Conclusions

Our present attempt to correlate the skin permeation rate (Log K_p) with theoretically calculated molecular descriptors, fragment descriptors, and external descriptor as Log *P* has led to successful QSAR models that relate this complex molecular property to structural characteristics of the molecules. Notably, all descriptors appearing in the multilinear equations (CODESSA PRO and ISIDA approach) and the ANN model have been derived from theoretical molecular calculations. The exception is the external descriptor Log *P*, for which the experimental data were taken from the literature. However, the computational methods were used and validated for the prediction of Log *P* values, used subsequently to QSAR models for skin permeation rate.

The current computational power available for chemical research allows carrying out such QSAR calculations for large data sets in a realistic time frame. The results obtained for this work indicate that both the multilinear regression and the ANN models exhibit good prediction capabilities. Thus, in principle, the QSAR models developed in our present work can be used for the prediction of skin permeation rate for additional compounds.

Acknowledgment. The Estonian Science Foundation grant No. 4548 is acknowledged for the partial support of this work.

Supporting Information Available: Log *P* calculated by three software programs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fujii, M.; Takeda, Y.; Yoshida, M.; Utoguchi, N.; Matsumoto, M.; Watanabe, Y. Comparison of skin permeation enhancement by 3-l-menthoxypropane-1,2-diol and l-menthol: the permeation of indomethacin and antipyrine through yucatan micropig skin and changes in infrared spectra and X-ray diffraction patterns of stratum corneum. *Int. J. Pharm.* **2003**, *258*, 217−223.

(2) Tanojo, H.; Boelsma, E.; Junginger, H. E.; Ponec, M.; Bodde, H. E. In vivo human skin permeability enhancement by oleic acid: a laser doppler velocimetry study. *J. Controlled Release.* **1999**, *58*, 97−104.

(3) Patel, H.; Berge, W.; Cronin, M. T. D. Quantitative structure−activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere* **2002**, *48*, 603−613.

(4) Schaefer, H.; Redelmeier, T. E. *Skin barrier: Principles of Percutaneoes Absorption*; S. Karger, Ed.; Zurich, Switzerland, 1996.

(5) Schnetz, E.; Fartasch, M. Microdialysis for the evaluation of penetration through the human skin barrier − a promising tool for future research? *Eur J. Pharm. Sci*. **2002**, *12*, 165−174.

(6) Tang, H.; Blankschtein, D.; Langer, R. Effects of low-Frequency ultrasound on the transdermal permeation of mannitol: comparative studies with in vivo and in vitro skin. *J. Pharm. Sci.* **2002**, *91*, 1776−1794.

(7) Cronin, M. T. D.; Dearden, J. C.; Gupta, R.; Moss, G. P. An investigation of the mechanism of flux Across polydimethylsiloxane membranes by use of quantitative structure-permeability relationships, *J. Pharm. Pharmacol.* **1999**, *50*, 143−152.

(8) Franz, J. T. Percutaneous absorption on the relevance of in vitro data. *J. Invest. Dermatol.* **1975**, *64*, 190−195.

(9) Bronaugh, R. L.; Stewart, R. F.; Congdon, E. R.; Giles, A. L. Methods for in vitro percutaneous absorption studies. I. Comparison with in vivo results. *Toxicol. Appl. Pharmacol.* **1982**, *62*, 474−480.

(10) Barry, B. W. Dermatological formulations. *Percutaneous absorption. (Drugs and the pharmaceutical sciences)*; Marcel Dekker: New York, 1983.

(11) Bronaugh, R. L.; Maibach, H. I. Percutaneous absorption of nitroaromatic compounds: in vivo and in vitro studies in the human and monkey. *J. Invest. Dermatol.* **1985**, *84*, 180−183.

(12) Friend, D. R In vitro skin permeation techniques. *J. Controlled Release.* **1992**, *18*, 235−248.

(13) Bartek, M. J.; LaBudde, J. A.; Maibach, H. I. Skin permeability in vivo. Comparison in rat, rabbit, pig and man. J. *Invest. Dermatol.* **1972**, *58*, 114−123.

(14) Chow, C.; Chow, A. Y. K.; Downie, R. H. H. S Buttar, Percutaneous absorption of hexachlorophene in rats, guinea pigs and pigs. *Toxicology* **1978**, *9*, 147−154.

(15) Bronaugh, R. L.; Stewart, R. F.; Congdon, E. R. Methods for in vitro percutaneous absorption studies. II. Animal models for human skin. *Toxicol. Appl. Pharmacol*. **1982**, *62*, 481−488.

(16) Itoh, T.; Xia, J.; Magavi, R.; Nishihata, T.; Rytting, J. H. Use of shed snake skin as a model membrane for in vitro percutaneous penetration studies: Comparison with human skin. *Pharm. Res.* **1990**, *7*, 1042−1047.

(17) Lin, S. Y.; Hou, S. J.; Hsu, T. H.; Yeh, F. L. Comparisons of different animal skins with human skin in drug percutaneous penetration studies. *Methods Find. Exp. Clin. Pharmacol*. **1992**, *14*, 645−654.

(18) Harada, K.; Murakami, T.; Kawasaki, E.; Higashi, Y.; Yamamoto, S.; Yata, N. In-vitro permeability to salicylic acid of human, rodent, and shed snake skin. *J. Pharm. Pharmacol*. **1993**, *45*, 414−418.

(19) Patel, H.; Cronin, M. Determination of the optimal physicochemical parameters to use in a QSAR-approach to predict skin permeation rate, Final Report CEFIC-LRI Project No. NMALRI-A2.2UNJM-0007, June 2001.

(20) Pugh, W. J.; Degim, I. T.; Hadgraft, J. Epidermal permeability-penetrant structure relationships: QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. *Int. J. Pharm*. **2000**, *197*, 203−211.

(21) Scheuplein, R. J.; Blank, I. H. Permeability of the skin. *Physiol. Rev*. **1971**, *51*, 702−747.

(22) Lien, E. J.; Tong, G. L. Physicochemical properties and percutaneous absorption of drugs. *J. Soc. Cosmet. Chem*. **1973**, *24*, 371−384.

(23) Roberts, M. S.; Anderson, R. A.; Swarbrick, J. Permeability of human epidermis to phenolic compounds. *J. Pharm. Pharmacol*. **1977**, *29*, 677−683.

(24) Flynn, G. L. Physicochemical determinants of skin absorption. In *Principles of route-to-route exptrapolation for risk assessment*; Gerrity, T. R., and Henry, C. J., Eds.; Elsevier: New York, 1990.

(25) Moss, G. P.; Dearden, J. C.; Patel, H.; Cronin, M. T. D. Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicol. in Vitro* **2002**, *16*, 299−317.

(26) ISIDA, May 2005. http://infochim.u-strasbg.fr/recherche/isida/index-.php (accessed May 2005).

(27) Wilschut, A.; ten Berge, W.; Robinson, P. J.; McKone, T. Estimating skin permeation. The validation of five mathematical skin permeation models. *Chemosphere* **1995**, *30*, 1275−1296.

(28) Johnson, M.; Blankschtein, D.; Langer, R. Evaluation of solute permeation through the stratum corneum: lateral bilayer diffusion as the primary transport mechanism. *J. Pharm. Sci.* **1997**, *86*, 1162−1172.

(29) Raevsky, O. A.; Schaper, K.-J. Quantitative estimation of hydrogen bond to permeability and absorption processes of some chemicals and drugs. *Eur. J. Med. Chem.* **1998**, *33*, 799−807.
(30) Nakai, J. S.; Stathopulos, P.; Campbell, G.; Chu, I.; Li-Muller, A.; Aucoin, R. Penetration of chloroform, trichloroethylene, and tetrachloroethylene through human skin. *J. Tox. Environ. Health* **1999**, *58*, 157−170.
(31) Buchwald, P.; Bodor, N. A simple, predictive, structure-based skin permeability model. *J. Pharm. Pharmacol*. **2001**, *53*, 1087−1098.
(32) Xu, X.; Mariano, T. M.; Laskin, J. D.; Weisel, C. P. Percutaneous absorption of trihalomethanes, haloacetic acids, and haloketones. *Toxicol. Appl. Pharmacol*. **2002**, *184*, 19−26.
(33) Kwon, K. H.; Kim, K. I.; Jun, W. J.; Shin, D. H.; Cho, H. Y.; Hong, B. S. In vitro and in vivo effects, of macrophage-stimulatory polysaccharide from leaves of perilla frutescens var. crispa. *Biol. Pharm. Bull.* **2002**, *25*, 367 − 371.
(34) Plessis, J.; Pugh, W. J.; Judefiend, A.; Hadgraft, J. Physico-chemical determinants of dermal drug delivery: effects of the number and substitution pattern of polar groups. *Eur. J. Pharm. Sci.* **2002**, *16*, 107−112.
(35) Abraham, M. H.; Martins, F. Human skin permeation and partition: general linear free-energy relationship analyses. *J. Pharmacol. Sci.* **2004**, *93*, 1508−1523.
(36) Diez Sales, O.; Castellano, A. L.; Lacer, F. J. M.; Dominiquez, M. H. An in vitro percutaneous absorption study of nonionic compounds across human skin. *Pharmazie* **1993**, *48*, 684−686.
(37) Interactive Analysis LogP and LogW Predictors, May 2005. http://www.logp.com/ (accessed May 2005).
(38) Syracuse Research Corporation, September 2004. http://esc.syrres-.com/interkow/kowdemo.htm (accessed May 2005).
(39) Daylight Chemical Information Systems, Inc May 2005. http://www.daylight.com/release/f_manuals.html (accessed − May 2005).
(40) Mdl Information Systems, Inc., May 2005. http://mdl.com (accessed May 2005).
(41) Hypercube, Inc., May 2005. http://www.hyper.com (accessed May 2005).
(42) Stewart, J. J. P. MOPAC 7.0, QCPE #455.
(43) CODESSA PRO, May 2005. http://www.codessa-pro.com (accessed May 205).
(44) Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley-Interscience: New York, 2000.
(45) Benson, S. W.; Buss, J. H. Additivity rules for estimation of thermo chemical properties. *J. Chem. Phys*. **1958**, *29*, 546−572.
(46) Haykin, S. *Neural Networks. A Comprehensive Foundation*, Pearson Education Inc, 1999.
(47) Stone, M. An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion. *J. R. Stat. Soc.* **1977**, *38*, 44−47.
(48) Dobchev, D. A.; Karelson, M. Reparametrized Austin Model for the Quantitative − Property Relationships in Liquid Media. *J. Mol. Model*. **2006**, *12*, 503−512.
(49) Solov'ev, V. P.; Varnek, A Anti-HIV Activity of HEPT, TIBO and Cyclic Urea Derivatives: Structure−Property Studies, Focused Combinatorial Library Generation and Hits Selection Using Substructural Molecular Fragments Method. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1703−1719.
(50) Varnek, A.; Fourches, D.; Solov'ev, V. P.; Baulin, V. E.; Turanov, A. N.; Karandashev, V. K.; Fara, D.; Katritzky, A. R. "In Silico" Design of New Uranyl Extractants Based on Phosphoryl-Containing Podands: QSPR Studies, Generation and Screening of Virtual Combinatorial Library and Experimental Tests. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1365−1382.