Acta Pharmacol Sin 2007 Apr; 28 (4): 591–600

Blackwell Publishing

Full-length article

Prediction of human skin permeability using artificial neural network (ANN) modeling[1]

Long-jian CHEN[2], Guo-ping LIAN[3], Lu-jia HAN[2,4]

[2]*College of Engineering, China Agricultural University, Beijing 100083, China;* [3]*Unilever Corporate Research, Bedford MK44 1LQ, UK*

Key words

ANN model; diffusion; permeability; quantitative structure-activity relationship; skin

[1] Project supported by grant-in-aid for scientific research from Unilever UK Central Resources Limited.
[4] Correspondence to Prof Lu-jia HAN.
Phn 86-10-6273-6313.
Fax 86-10-6273-6778.
E-mail clj1020@googlemail.com

Received 2006-09-14
Accepted 2006-10-26

doi: 10.1111/j.1745-7254.2007.00528.x

Abstract

Aim: To develop an artificial neural network (ANN) model for predicting skin permeability (log K_p) of new chemical entities. **Methods**: A large dataset of 215 experimental data points was compiled from the literature. The dataset was subdivided into 5 subsets and 4 of them were used to train and validate an ANN model. The same 4 datasets were also used to build a multiple linear regression (MLR) model. The remaining dataset was then used to test the 2 models. Abraham descriptors were employed as inputs into the 2 models. Model predictions were compared with the experimental results. In addition, the relationship between log K_p and Abraham descriptors were investigated. **Results:** The regression results of the MLR model were n=215, determination coefficient (R^2)=0.699, mean square error (MSE)=0.243, and F=493.556. The ANN model gave improved results with n=215, R^2=0.832, MSE=0.136, and F=1050.653. The ANN model suggests that the relationship between log K_p and Abraham descriptors is non-linear. **Conclusion:** The study suggests that Abraham descriptors may be used to predict skin permeability, and the ANN model gives improved prediction of skin permeability.

Introduction

Skin permeability is of relevance to a number of applications including the design of skin cream products, risk assessment of hazardous chemicals, and in particular, transdermal delivery of drugs. There has been a continuous interest in predicting skin permeability since Scheuplein[1,2] and Scheuplein and Blank[3,4] first introduced anatomically-based physicochemical models describing percutaneous absorption. This is partly due to ethical difficulties with respect to human and animal experiments and partly due to economic considerations and increasing legislation in the risk assessment of industrial chemicals, for example, the newly-proposed European chemicals strategy: Registration, Evaluation, Authorization and Restriction of Chemicals (REACH)[5].

Previous studies on predicting skin permeability can be categorized into mechanistic and empirical models[3,6,7]. The advantage of many mechanistic models is that they can provide insights into the mechanisms of skin permeation. However, most mechanistic models are complex in nature and several challenges remain for practical use. A main challenge is that mechanistic models involve parameters that are not easily measurable and attainable. Assumptions also have to be made. Some assumptions may be oversimplified and not necessary apply to real situation. In the domain of skin permeability prediction, empirical models have been also frequently reported[8]. Many empirical models employ the so-called quantitative structure-activity relationship (QSAR) methods which attempt to relate statistically the biological activity of a series of compounds to their physicochemical and/or structural properties[9]. QSAR methods, stretching back over a century, had been applied in many fields[10,11]. In the last 30 years, QSAR methods had been developed to relate percutaneous penetration properties of a range of chemical compounds to their physicochemical parameters. Compared with the mechanism model, the QSAR model for skin permeability does not consider the dynamic process of skin permeation. Most QSAR models for skin permeability work well within the range of the experimental data, but often can not be extrapolated. From the point of view of practical application, the approach is simple and can provide accept-

able predictions once validated. Various QSAR models for skin permeability have been reported. Moss *et al*[8] recently gave a comprehensive review on different QSAR models for skin permeability. An early study by Potts and Guy[12] analyzed the data of Flynn[13] and related the skin permeability of the analyzed compounds to their octanol-water partition property and molecular weight. They proposed the empirical equation of skin permeability as $\log K_p$=0.71$\log K_{ow}$ -0.0061MW-6.3, where $\log K_p$ is skin permeability is given in cm/s, $\log K_{ow}$ is solute partition coefficient in octanol/water, MW is molecular weight. Different empirical equations were also proposed by other researchers using other structural molecular parameters such as the number of hydrogen bonds and molecular volume[14,15]. Most QSAR models for skin permeability employed the multiple linear regression (MLR) method. The method provides an efficient way to determine the most relevant physicochemical descriptors. The main drawback of the MLR approach is that it assumes linearity between the descriptors and skin permeability. The artificial neural network (ANN) is similar to the MLR approach, but is more suitable for extracting both linear and nonlinear effects of chosen descriptors on skin permeability. Recently, Degim and colleagues[16,17] used the ANN model to predict skin permeability. The dataset was limited to 40 chemical compounds in the range of -0.77≤$\log K_{ow}$≤4.57 and 32 Da≤MW≤ 389 Da. The descriptors used include the partial charge, $\log K_{ow}$ and MW of each compound. Abraham *et al*[18] argued that $\log K_{ow}$ was an empirical colligation variable and did not give the actual structural features of the chemical compounds that influence skin permeability. Therefore, some researchers have related skin permeability to Abraham descriptors which they believe can better describe the actual features of molecules and improve the precision of the model[19].

In this paper, we report that the ANN model predicts skin permeability of chemicals using Abraham descriptors. A large database of skin permeability containing 215 data points was compiled from literature. The correlation between the skin permeability coefficient and the Abraham descriptors were obtained from the trained neural network. In addition, the predictability of the neural network model was compared to the MLR model. The ANN model was shown to give better prediction results which indicate non-linearity and complexity of correlation between Abraham descriptors and skin permeability. Some insight into the degree of nonlinear behavior of every Abraham descriptors has also been assessed with a functional dependence to understand relationships.

Materials and methods

Data set A structurally-diverse set of compounds was selected to construct the MLR and ANN models (n=215, -2.11≤$\log K_{ow}$≤7.6, 18 Da≤MW≤518 Da, -0.85≤$\log K_p$(cm/h) ≤ -5.22). Skin permeability data expressed in $\log K_p$(cm/h), were obtained from literature and regulatory sources[13,20,21]. Abraham descriptors of R_2 (excess molar refraction), π_2^H (the dipolarity/polarizability), $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$ (the overall or effective hydrogen-bond acidity and basicity), and V_x (the McGowan characteristic volume) were obtained using Abraham Solvation Parameters (ABSOLV) program (Pharma Algorithms Software, Toronto, Canada). The program was written to read molecular structures as Simplified Molecular Input Line Entry System (SMILES) strings which were obtained from the PubChem net database[22]. The dataset is listed in Table 1.

Prediction models The chemical compounds were listed alphabetically according to their chemical names and divided into 5 subsets. Details about the division of the sub-datasets are also shown in Table 1. Subsets 1, 3, and 5 were used as the training dataset, subset 4 as the validation dataset, and subset 2 as the testing dataset. For comparison, a MLR model had also been generated using subsets 1, 3, 4, and 5. The Neural Network Toolbox of MATLAB 7.0.2 was used to construct the ANN model. Back-propagation networks had been employed in this study. There are some empirical rules about constructing artificial neural networks, for example, 2 hidden layers are sufficient for generality[23]. The number of units of the hidden layer normally is not more than twice the input [24]. In our example, there were 5 inputs corresponding to the 5 Abraham descriptors and 1 output for skin permeability. To determine the ANN architecture, all network structures with the maximum allowed numbers of hidden layers and hidden units limited to 2 and 10, respectively, had been investigated. Each network was trained by 3-folding cross validation using the 3 subsets 1, 3, and 5. For the training of each network, the average mean square errors (MSE) between the model prediction and experimental data were obtained. The ANN architecture with the smallest average value of MSE was chosen. The MSE is defined as follows:

$$MSE = \frac{\sum_{i=1}^{N} (\log K_{p,obs} - \log K_{p,cal})^2}{N} \quad (1)$$

where $\log K_{p,obs}$ is the experimental value, $\log K_{p,cal}$ is the predicted value, and N is the number of data points in the training dataset.

After determining the ANN architecture, the 3 subsets of 1, 3, and 5 were combined with subset 4 to further tune the ANN parameters to obtain the final ANN model. Finally, subset 2 that had not been used for training the network was used to test the predictability of the model. The same subset

Table 1. Names, R_2, π_2^H, $\Sigma\alpha_2^H$, $\Sigma\beta_2^H$, and V_x of the compounds used in the study.

Name	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_x	log K_p (cm/h) Experimental	log K_p (cm/h) MLR	log K_p (cm/h) ANN-5711	Subset
1,1,1-trichloroethane	0.3100	0.4400	0.0000	0.0100	0.7576	-2.3500	-1.6836	-2.0838	1
1,3-Dichloropropene	0.4000	0.5600	0.0000	0.1000	0.7331	-2.0000	-1.9982	-2.3546	2
2,3-Butanediol	0.4100	0.7300	0.6300	0.5900	0.7896	-4.3900	-2.7314	-3.4952	3
2,3-Butanediol	0.4100	0.7300	0.6300	0.5900	0.7896	-4.3010	-2.7314	-3.4952	4
2,4,6-Trichlorophenol	1.0100	0.8000	0.6800	0.1500	1.1423	-1.2262	-0.9002	-1.3246	5
2,4-Dichlorophenol	0.9600	0.8400	0.5300	0.1900	1.0199	-1.2211	-1.3610	-1.3499	1
2-Amino-4-nitrophenol	1.3200	1.7800	0.7500	0.6400	1.0491	-3.1810	-2.8071	-2.9771	2
2-Butanone	0.2200	0.6700	0.0000	0.3400	0.6879	-2.9500	-2.7818	-2.9052	3
2-Butanone	0.2200	0.6700	0.0000	0.3400	0.6879	-2.3468	-2.7818	-2.9052	4
2-Butoxyethanol	0.2800	0.5600	0.2300	0.6300	1.0714	-2.8500	-2.5123	-2.6511	5
2-Chlorophenol	0.8530	0.8800	0.3200	0.3100	0.8975	-1.4802	-2.0682	-1.6381	1
2-Chlorophenol	0.8530	0.8800	0.3200	0.3100	0.8975	-1.4401	-2.0682	-1.6381	2
2-Ethoxyethanol	0.2800	0.5500	0.2300	0.6200	0.7896	-3.6021	-2.9810	-3.2241	3
2-Heptanone	0.2100	0.6900	0.0000	0.3500	1.1106	-2.0000	-2.0744	-2.3402	4
2-Hexanone	0.2100	0.6800	0.0000	0.3500	0.9697	-2.3500	-2.3165	-2.6203	5
2-Isopropyl-5-methylphenol	0.8220	0.7900	0.5200	0.4400	1.3387	-1.2737	-1.3916	-1.2267	1
2-Methyphenol	0.8400	0.8600	0.5200	0.3000	0.9160	-1.8037	-1.8638	-1.614	2
2-Naphthol	1.5200	1.0800	0.6100	0.4000	1.1441	-1.5544	-1.5691	-1.5667	3
2-Naphthol	1.5200	1.0800	0.6100	0.4000	1.1441	-1.5918	-1.5691	-1.5667	4
2-Nitro-4-phenylenediamine	1.4700	1.8600	0.4700	0.7900	1.0902	-3.3010	-3.2859	-3.6195	5
2-Nitrophenol	1.0150	1.0500	0.0500	0.3700	0.9493	-1.0037	-2.3776	-1.4894	1
2-Pentanone	0.2100	0.6800	0.0000	0.3400	0.8288	-2.6000	-2.5431	-2.8097	2
2-Phenylenediamine	1.1900	1.4200	0.4700	0.7100	0.9160	-3.3500	-3.1642	-3.1595	3
2-Phenylethanol	0.8110	0.9100	0.3000	0.6400	1.0569	-1.8861	-2.6033	-1.8564	4
3,4-Dimethylphenol	0.8300	0.8600	0.5600	0.3900	1.0569	-1.4437	-1.7989	-1.4369	5
3,4-Xylenol	0.8300	0.7900	0.5000	0.3900	1.0569	-1.4437	-1.7861	-1.4038	1
3-Cresol	0.8300	0.7900	0.5000	0.3900	1.0569	-1.8200	-1.7861	-1.4038	2
3-Methyphenol	0.8220	0.8800	0.5700	0.3400	0.9160	-1.8137	-1.9439	-1.7306	3
3-Nitrophenol	1.0500	1.5700	0.7900	0.2300	0.9493	-2.2534	-1.9335	-2.28	4
4-Amino-2-nitrophenol	1.2600	1.6600	0.4100	0.6600	1.0491	-2.5500	-3.0148	-2.8256	5
4-Bromophenol	1.0800	1.1700	0.6700	0.2000	0.9501	-1.4425	-1.6259	-1.6509	1
4-Chloro-3,5-dimethylphenol	0.9250	0.9600	0.6400	0.2100	1.1793	-1.2337	-1.1538	-1.2474	2
4-Chloro-3,5-xylenol	0.9800	0.9000	0.6700	0.3800	1.1793	-1.2800	-1.4631	-1.3089	3
4-Chloro-3-methylphenol	0.9200	1.0200	0.6500	0.2200	1.0384	-1.2637	-1.4680	-1.4167	4
4-Chlorophenol	0.9150	1.0800	0.6700	0.2000	0.8975	-1.4400	-1.7056	-1.8304	5
4-Cresol	0.8100	0.8500	0.5000	0.3900	0.9160	-1.7600	-2.0890	-1.7446	1
4-Cyanophenol	0.9600	1.3900	0.6700	0.5300	0.9298	-1.9830	-2.6390	-2.6994	2
4-Ethylphenol	0.8000	0.9000	0.5500	0.3600	1.0569	-1.4600	-1.7771	-1.4498	3
4-Methyl-2-pentanol	0.2200	0.4300	0.3100	0.3700	1.0127	-2.3300	-1.8774	-1.5902	4
4-Methyphenol	0.8200	0.8700	0.5200	0.3100	0.9160	-1.7537	-1.9015	-1.6551	5
4-Nitrophenol	1.0700	1.7200	0.8200	0.2600	0.9493	-2.2487	-2.0916	-2.3413	1
4-Phenylenediamine	1.2900	1.5900	0.6700	0.9800	0.9747	-3.6200	-3.6473	-4.3006	2
Acetaldehyde	0.2400	0.7200	0.0000	0.3200	0.4061	-3.1500	-3.2668	-2.9687	3
Acetic acid	0.2650	0.6500	0.6100	0.4400	0.4648	-3.2100	-2.9599	-3.0437	4
Acetone	0.2200	0.6700	0.0000	0.3400	0.5470	-3.2900	-3.0316	-2.934	5
Acetonitrile	0.1900	0.7200	0.0000	0.2000	0.4042	-3.2100	-3.0084	-3.0054	1
Acrolein	0.3600	0.7800	0.0000	0.3700	0.5040	-3.0700	-3.2150	-3.1373	2
Acrylic acid	0.3000	0.6800	0.5700	0.4100	0.5627	-3.0500	-2.7554	-2.9288	3
Acrylonitrile	0.3100	0.7800	0.0000	0.2600	0.5021	-2.8700	-2.9798	-3.1657	4
Aldosterone	2.0100	3.4700	0.4000	1.9000	2.6890	-4.3010	-4.1378	-4.1964	5
Aldosterone	2.0100	3.4700	0.4000	1.9000	2.6890	-4.2366	-4.1378	-4.1964	1
Allyl alcohol	0.3300	0.5100	0.3100	0.3600	0.5470	-2.9500	-2.7043	-2.6116	2

Name	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_x	log K_p (cm/h) Experimental	log K_p (cm/h) MLR	log K_p (cm/h) ANN-5711	Subset
Amylobarbital	0.9800	1.3500	0.5200	1.2400	1.7966	-2.6440	-2.8169	-2.6831	3
Aniline	0.8600	1.0800	0.2300	0.4300	0.8162	-2.6500	-2.7050	-2.5977	4
Anisole	0.6200	0.7900	0.0000	0.3300	0.9160	-1.6000	-2.3112	-1.9517	5
Benzene	0.6100	0.5200	0.0000	0.1400	0.7164	-0.9547	-2.0187	-1.5637	1
Benzyl alcohol	0.8030	0.8700	0.3900	0.5600	0.9160	-2.2200	-2.5773	-2.0727	2
β-Naphtol	1.5000	1.2300	0.5000	0.4500	1.1441	-1.5600	-1.8834	-1.4807	3
Butanoic acid	0.2100	0.6200	0.6000	0.4500	0.7466	-2.9037	-2.4859	-2.9299	4
Butanone	0.1660	0.7000	0.0000	0.5100	0.6879	-2.3437	-3.2182	-2.6793	5
Butobarbital	0.9800	1.3600	0.5200	1.2100	1.6557	-3.7144	-3.0047	-3.1058	1
Butyl acrylate	0.1900	0.6500	0.0000	0.4200	1.1263	-2.0000	-2.1851	-2.2135	2
Butyric acid	0.1700	0.6200	0.5700	0.3600	0.7466	-3.0000	-2.3110	-2.602	3
Caffeine	1.9400	1.8100	0.0000	1.4700	1.3600	-4.0000	-4.5122	-4.2814	4
Catechol	0.9700	1.2300	0.5900	0.7400	0.8925	-2.7700	-3.1210	-3.4035	5
Chlorocresol	0.9600	0.9600	0.6700	0.3800	1.0384	-1.2600	-1.7660	-1.4581	1
Chloroxylenol	0.9800	0.9000	0.6700	0.3800	1.1793	-1.2774	-1.4631	-1.3089	2
Chloroxylenol	0.9800	0.9000	0.6700	0.3800	1.1793	-1.2300	-1.4631	-1.3089	3
Chlorpheniramine	1.5200	1.4900	0.0000	1.0200	2.2098	-2.6600	-1.8555	-2.2046	4
Codeine	2.0200	1.7800	0.2600	1.7500	2.2100	-4.3098	-3.4267	-4.1422	5
Cortexolone	1.9100	3.4500	0.3600	1.6000	2.7389	-4.1300	-3.3983	-3.865	1
Cortexone	1.6900	2.5500	0.1700	1.3200	2.6802	-3.3500	-2.3623	-2.3315	2
Corticosterone	1.8600	3.4300	0.4000	1.6300	2.7389	-3.5229	-3.4420	-3.8399	3
Corticosterone	1.8600	3.4300	0.4000	1.6300	2.7389	-3.2565	-3.4420	-3.8399	4
Corticosterone	1.8600	3.4300	0.4000	1.6300	2.7389	-4.0000	-3.4420	-3.8399	5
Cortisone	1.9600	3.5000	0.3600	1.8700	2.7546	-5.0000	-4.0193	-4.1684	1
Cumene	0.5900	0.6300	0.0000	0.1500	1.1391	-0.8500	-1.3842	-1.3346	2
Cyclohexanone	0.4200	0.7700	0.0000	0.3200	0.8611	-2.7400	-2.4377	-2.7939	3
Decanol	0.1910	0.4200	0.3700	0.4800	1.5763	-1.0970	-1.0946	-1.1635	4
Deoxycorticosterone	1.7400	3.5000	0.1400	1.3100	2.6802	-3.3437	-3.0761	-3.1987	5
Dexamethasone	2.0400	3.5100	0.7100	1.9200	2.9132	-4.1938	-3.5944	-3.453	1
Diclofenac	1.9700	1.8800	0.7800	0.8700	2.0300	-1.7399	-1.4385	-2.0849	2
Diethanolamine	0.5800	0.8200	0.6400	1.1900	0.8894	-4.3800	-3.9532	-4.3373	3
Diethylamine	0.1540	0.3500	0.1300	0.4800	0.7720	-2.7500	-2.6440	-2.282	4
Diethylcarbamazine	0.8200	1.4100	0.0000	1.3100	1.7241	-3.8900	-3.5658	-3.1854	5
Diethylether	0.0410	0.2500	0.0000	0.4500	0.7309	-1.7937	-2.6976	-2.3805	1
Dimethyl acetamide	0.3300	1.0600	0.0000	0.6400	0.7877	-2.8000	-3.5656	-2.7468	2
Dioxane	0.2900	0.5800	0.0000	0.4400	0.6810	-3.4500	-2.9332	-2.8405	3
Ephedrine	0.9800	0.9400	0.3800	1.1200	1.4385	-2.2218	-2.9527	-3.0943	4
Epichlorohydrin	0.3900	0.5500	0.0000	0.2500	0.6038	-3.4300	-2.5716	-2.8646	5
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.5229	-2.4102	-2.3908	1
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.4949	-2.4102	-2.3908	2
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.4685	-2.4102	-2.3908	3
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.4101	-2.4102	-2.3908	4
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.3979	-2.4102	-2.3908	5
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.3872	-2.4102	-2.3908	1
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.2840	-2.4102	-2.3908	2
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.2676	-2.4102	-2.3908	3
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.2147	-2.4102	-2.3908	4
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.4559	-2.4102	-2.3908	5
Estradiol	1.8000	3.3000	0.8800	0.9500	2.1988	-2.3809	-2.4102	-2.3908	1
Estratriol	2.0000	3.3600	1.4000	1.2200	2.2575	-3.5237	-2.5546	-4.1752	2
Estrone	1.7300	3.1000	0.5600	0.9100	2.1558	-2.4400	-2.4827	-2.762	3
Ethanol	0.2460	0.4200	0.3700	0.4800	0.4491	-3.0970	-3.0744	-3.178	4
Ethanol	0.2460	0.4200	0.3700	0.4800	0.4491	-3.0000	-3.0744	-3.178	5
Ethanol amine	0.4200	0.7200	0.4600	0.9400	0.5489	-4.0200	-4.0769	-4.1978	1

Name	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_x	log K_p (cm/h) Experimental	log K_p (cm/h) MLR	log K_p (cm/h) ANN-5711	Subset
Ethyl acrylate	0.1900	0.6400	0.0000	0.4200	0.8445	-2.3900	-2.6770	-2.6835	2
Ethyl benzene	0.5800	0.6400	0.0000	0.1200	0.9982	-1.1500	-1.5754	-1.4333	3
Ethyl ether	0.0400	0.3400	0.0000	0.2500	0.7309	-2.8000	-2.3028	-2.5462	4
Ethyl formate	0.0900	0.6300	0.0000	0.3400	0.6057	-3.0100	-2.9408	-2.7153	5
Ethylamine	0.2100	0.4900	0.2100	0.5700	0.4902	-3.0900	-3.3868	-3.2045	1
Ethylene dichloride	0.3800	0.4800	0.0000	0.1000	0.6352	-2.0000	-2.1168	-2.4551	2
Ethylene glycol	0.4100	0.7100	0.5400	0.5800	0.5078	-4.0700	-3.2542	-3.3893	3
Ethylhexyl phthalate	0.7200	1.4100	0.0000	0.8800	3.4014	-1.5200	0.3730	-2.0264	4
Fentanyl	1.8600	2.1800	0.0000	1.3300	2.8399	-2.5230	-1.8759	-2.3113	5
Fentanyl	1.8600	2.1800	0.0000	1.3300	2.8399	-2.2518	-1.8759	-2.3113	1
Formaldehyde	0.2100	0.6800	0.0000	0.3000	0.2652	-2.6500	-3.4495	-2.9194	2
Formic acid	0.3000	0.7900	0.7200	0.3400	0.3239	-3.5237	-2.9981	-3.0246	3
Heptanoic acid	0.1490	0.6000	0.6000	0.4500	1.1693	-1.7000	-1.7417	-2.1971	4
Hexachloroethane	0.8600	0.7400	0.0000	0.1500	1.1248	-1.4000	-1.4029	-1.2827	5
Hexanoic acid	0.1740	0.6000	0.6000	0.4500	1.0284	-1.8500	-1.9830	-2.5155	1
Hydrocortisone	2.0300	3.4900	0.7100	1.9000	2.7976	-3.6383	-3.7409	-3.9264	2
Hydrocortisone hexanoate	1.8100	3.0200	0.4600	2.1600	3.6587	-1.7447	-2.7016	-1.8399	3
Hydrocortisone octanoate	1.7700	3.0500	0.4600	2.1600	3.9404	-1.2076	-2.2389	-1.6749	4
Hydrocortisone hemipimelate	2.0200	3.5900	1.0600	2.6100	3.8740	-2.7447	-3.3219	-3.5623	5
Hydrocortisone hemisuccinate	2.1000	3.1500	1.0600	2.6100	3.4513	-3.2007	-3.7046	-3.4533	1
Hydrocortisone hydroxyhexanoate	2.0300	3.4900	0.8300	2.6400	3.7174	-3.0410	-3.7467	-3.4871	2
Hydrocortisone methylpimelate	1.9300	3.4800	0.4600	2.6100	4.0149	-2.2676	-3.4299	-2.3344	3
Hydrocortisone methylsuccinate	1.9900	3.1000	0.4600	2.6100	3.5922	-3.6778	-3.8656	-3.2813	4
Hydrocortisone *N*,*N*-dimethylsuccinamate	2.2100	3.7500	0.4800	2.8600	3.7742	-4.1739	-4.5372	-3.5171	5
Hydrocortisone pimelamate	2.2100	3.9000	0.9600	2.8400	3.9151	-3.0506	-4.0269	-3.7949	1
Hydrocortisone proprinate	1.8700	2.9000	0.4600	2.1600	3.2360	-2.4685	-3.3381	-2.6183	2
Hydrocortisone succinamate	2.3100	3.3500	1.0000	2.8400	3.4924	-4.5850	-4.2904	-3.8825	3
Hydromorphone	2.0400	1.7900	0.2700	1.3200	2.0648	-4.0757	-2.6793	-3.6456	4
Hydroxypregnenolone	1.5500	3.3500	0.5700	1.3500	2.7232	-3.2200	-2.7460	-3.6066	5
Hydroxyprogesterone	1.6400	3.3500	0.2500	1.3100	2.6802	-3.2200	-2.9186	-3.3051	1
Indomethacin	2.3900	2.7200	0.5900	1.1900	2.5300	-1.8297	-1.9320	-1.9904	2
Isoamyl alcohol	0.2100	0.4400	0.3100	0.3400	0.8718	-2.0000	-2.0686	-1.9281	3
Isobutyl alcohol	0.2100	0.4400	0.3100	0.3400	0.7309	-2.6500	-2.3185	-2.2298	4
Isopropyl alcohol	0.2200	0.4300	0.3100	0.3400	0.5900	-3.0500	-2.5572	-2.4799	5
Isopropylamine	0.2100	0.4800	0.2100	0.6000	0.6311	-2.9000	-3.1990	-3.3561	1
Isoquinoline	1.3200	1.1500	0.0000	0.4600	1.0443	-1.7747	-2.4264	-1.2925	2
Lidocaine	1.1000	1.5000	0.2600	1.1700	2.0589	-2.4012	-2.4430	-2.1216	3
Lignocaine	1.0100	1.4900	0.1100	1.2700	2.0589	-2.4037	-2.8011	-2.2476	4
m-Cresol	0.8200	0.8800	0.5700	0.3400	0.9160	-1.8182	-1.9445	-1.733	5
Meperidine	0.9900	1.2600	0.0000	0.9700	2.0501	-2.4300	-2.0246	-2.252	1
Methanol	0.2780	0.4400	0.4300	0.4700	0.3082	-3.3010	-3.2644	-3.2006	2
Methyl 4-hydroxybenzoate	0.9000	1.3700	0.6900	0.4500	1.1313	-2.0400	-2.0871	-2.0484	3
Methyl acrylate	0.1900	0.6400	0.0000	0.4100	0.7036	-2.6800	-2.9036	-2.7817	4
Methyl acrylic acid	0.1900	0.6400	0.0000	0.4100	0.7036	-2.5800	-2.9036	-2.7817	5
Methylhydroxybenzoate	0.8000	1.0600	0.1300	0.4000	1.1313	-2.0400	-2.1502	-1.7291	1
m-Nitrophenol	1.0500	1.4700	0.6900	0.4900	0.9493	-2.2490	-2.5290	-2.6225	2
Monomethylhydrazine	0.4100	0.5500	0.3400	0.9000	0.4491	-3.7500	-4.1155	-4.2119	3
Morphine	2.1000	1.6800	0.5500	1.7600	2.0600	-4.2518	-3.4123	-4.1715	4
Morpholine	0.3700	0.6200	0.1600	0.6600	0.7221	-3.8600	-3.2652	-3.3589	5
N,*N*-Dimethyl aniline	0.7900	0.9100	0.0000	0.4600	1.0980	-1.7000	-2.3255	-1.4955	1
Naproxen	1.6200	1.4000	0.5900	0.7500	1.7800	-2.5376	-1.4816	-1.7984	2
n-Butanol	0.2240	0.4200	0.3700	0.4800	0.7309	-2.6576	-2.5823	-2.7753	3
n-Butanol	0.2240	0.4200	0.3700	0.4800	0.7309	-2.6021	-2.5823	-2.7753	4

Name	R_2	π_2^H	$\Sigma\alpha_2^H$	$\Sigma\beta_2^H$	V_x	log K_p (cm/h) Experimental	log K_p (cm/h) MLR	log K_p (cm/h) ANN-5711	Subset
n-Butanol	0.2240	0.4200	0.3700	0.4800	0.7309	-2.5229	-2.5823	-2.7753	5
n-Heptanol	0.2110	0.4200	0.3700	0.4800	1.1536	-1.4950	-1.8372	-1.5464	1
n-Heptanol	0.2110	0.4200	0.3700	0.4800	1.1536	-1.4248	-1.8372	-1.5464	2
n-Hexanol	0.2100	0.4200	0.3700	0.4800	1.0127	-1.8861	-2.0874	-1.9345	3
n-Hexanol	0.2100	0.4200	0.3700	0.4800	1.0127	-1.5575	-2.0874	-1.9345	4
n-Hexanol	0.2100	0.4200	0.3700	0.4800	1.0127	-1.6990	-2.0874	-1.9345	5
Nicotine	1.0500	1.0900	0.0000	1.1100	1.3700	-2.4815	-3.4041	-2.7985	1
n-Nitrosodiethanolamine	0.8000	1.4600	0.4800	1.2400	1.0049	-5.2200	-4.3940	-4.297	2
n-Nonanol	0.1930	0.4200	0.3700	0.4800	1.4354	-1.2218	-1.3437	-1.2174	3
n-Octanol	0.1990	0.4200	0.3700	0.4800	1.2950	-1.2800	-1.5906	-1.3315	4
n-Octanol	0.1990	0.4200	0.3700	0.4800	1.2950	-0.9586	-1.5906	-1.3315	5
n-Propanol	0.2360	0.4200	0.3700	0.4800	0.5900	-2.9208	-2.8280	-3.0343	1
n-Propanol	0.2360	0.4200	0.3700	0.4800	0.5900	-2.7700	-2.8280	-3.0343	2
o-Chlorophenol	0.8400	0.8400	0.3300	0.3000	0.8975	-1.4820	-2.0116	-1.5816	3
o-Cresol	0.8400	0.8600	0.5200	0.3000	0.9160	-1.8050	-1.8638	-1.614	4
Octanoic acid	0.1500	0.6000	0.6000	0.4500	1.3102	-1.6000	-1.4915	-1.8222	5
o-Phenylenediamine	1.1900	1.4200	0.4700	0.7100	0.9160	-3.3468	-3.1642	-3.1595	1
p-Cresol	0.8200	0.8700	0.5700	0.3100	0.9160	-1.7570	-1.8671	-1.6743	2
Pentanoic acid	0.2050	0.6000	0.6000	0.4500	0.8875	-2.7000	-2.2223	-2.7732	3
Pentanol	0.2190	0.4200	0.3700	0.4800	0.8718	-2.2220	-2.3341	-2.3789	4
p-ethylphenol	0.8100	0.8500	0.5000	0.3900	1.0569	-1.4580	-1.8392	-1.426	5
Phenobarbital	1.5600	1.8100	0.5200	1.2900	1.6999	-3.3439	-3.2630	-3.523	1
Phenol	0.8050	0.8900	0.6000	0.3000	0.7751	-2.0851	-2.0936	-2.2064	2
Phenol	0.8050	0.8900	0.6000	0.3000	0.7751	-1.8861	-2.0936	-2.2064	3
Phenylglycidyl ether	0.8800	1.0200	0.0000	0.5500	1.1479	-2.8400	-2.5005	-1.488	4
p-Naphthol	1.5000	1.2300	0.5000	0.4500	1.1441	-1.5544	-1.8834	-1.4807	5
p-Nitrophenol	1.0500	1.4700	0.6700	0.4900	0.9493	-2.2540	-2.5428	-2.6078	1
p-Phenylenediamine	1.1700	1.5000	0.4500	0.7400	0.9160	-3.6198	-3.3162	-3.3914	2
Pregnenolone	1.3600	3.2900	0.3200	1.1800	2.6645	-2.8200	-2.6450	-2.6836	3
Progesterone	1.4500	3.2900	0.0000	1.1400	2.6215	-1.8861	-2.8176	-3.2079	4
Propanol	0.2360	0.4200	0.3700	0.4800	0.5900	-2.8540	-2.8280	-3.0343	5
Propionic acid	0.2330	0.6500	0.6000	0.4500	0.6057	-2.9400	-2.7511	-3.0048	1
Propylene dichloride	0.3800	0.4700	0.0000	0.1300	0.7761	-2.0000	-1.9290	-1.9722	2
Propylene oxide	0.2600	0.4000	0.0000	0.2500	0.4814	-3.0500	-2.7169	-2.9708	3
Pyridine	0.6000	0.8200	0.0000	0.4000	0.6753	-2.7400	-2.9305	-2.9693	4
Resorcinol	0.9800	1.0000	1.1000	0.5800	0.8338	-2.8200	-2.3221	-3.2217	5
Salicylic acid	0.9100	1.1000	0.7000	0.4000	0.9904	-1.9031	-2.0020	-1.9239	1
Salicylic acid	0.9100	1.1000	0.7000	0.4000	0.9904	-1.5171	-2.0020	-1.9239	2
Scopolamine	1.6400	1.9600	0.3500	1.8400	2.2300	-4.3010	-3.8062	-3.9703	3
Sufentanil	1.8100	2.2200	0.0000	1.5000	3.1051	-2.2600	-1.8484	-2.3196	4
Sufentanyl	1.8100	2.2200	0.0000	1.5000	3.1051	-1.9208	-1.8484	-2.3196	5
Testosterone	1.5400	2.5900	0.3200	1.1900	2.3827	-2.6576	-2.5664	-2.5048	1
Testosterone	1.5400	2.5900	0.3200	1.1900	2.3827	-2.2708	-2.5664	-2.5048	2
Thymol	0.8400	0.7800	0.5000	0.4200	1.3387	-1.2596	-1.3450	-1.2362	3
Toluene	0.5800	0.6300	0.0000	0.1200	0.8573	-1.3000	-1.8175	-1.6005	4
Triethylamine	0.1700	0.3700	0.0000	0.5300	1.0538	-2.3100	-2.3598	-1.8299	5
Urea	0.6300	1.1700	0.7200	0.6900	0.4648	-3.8300	-3.7428	-3.7431	1
Vinyl acetate	0.1900	0.6400	0.0000	0.4100	0.7036	-2.7300	-2.9036	-2.7817	2
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-3.3010	-3.0695	-3.0889	3
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.8827	-3.0695	-3.0889	4
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.8539	-3.0695	-3.0889	5
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.8069	-3.0695	-3.0889	1
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.8013	-3.0695	-3.0889	2
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.7670	-3.0695	-3.0889	3
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-3.5229	-3.0695	-3.0889	4
Water	0.0000	0.4500	0.8200	0.3500	0.1673	-2.8125	-3.0695	-3.0889	5

2 was also used to test the MLR model. The early-stopping method was used for the training and validation.

Results

Comparison between the MLR and ANN models The best ANN structure with minimal MSE contained 7 neurons at the first hidden layer and 1 neuron at the second hidden layer. The model was described as ANN-5711, which corresponded to a minimal MSE of 0.10511. Figure 1 shows the relationship between predicted values and experimental results. The statistical test results for the ANN model are shown in Table 2.

Many previous QSAR models for skin permeability were based on the MLR method. Here, for comparison, subsets 1, 3, 4, and 5 were also used to build a MLR model using the 5 Abraham descriptors. The best-fitted equation was as follows:

$$\log K_p\,(\mathrm{cm/h}) = -2.769 + 0.339R_2 - 0.772\pi_2^H + 0.687\sum\alpha_2^H - 2.323\sum\beta_2^H + 1.773V_x \quad (2)$$

When using the above equation, the *MSE* for the predicted log K_P was found to be 0.25 for datasets 1, 3, 4, and 5, and 0.21 for the testing dataset 2. Figure 2 shows the relationship between the experimental results and predicted values using the MLR model of equation 2. The statistical test results for the MLR model are shown in Table 3.

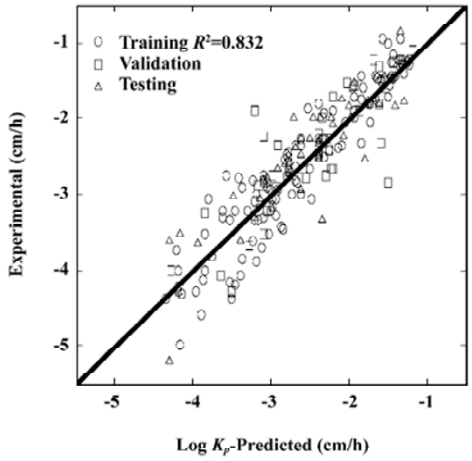

Figure 1. Comparison of experimental data with predictions made with ANN model. Line represents 1 to 1 correspondence line.

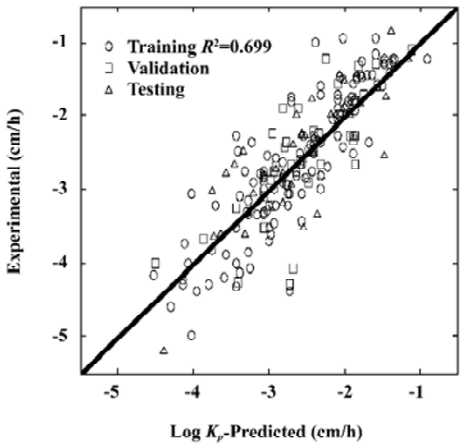

Figure 2. Comparison of experimental data with predictions made with MLR model. Line represents 1 to 1 correspondence line.

Table 2. The statistical test results of ANN model.

Regression statistics					
R	0.912				
R Square	0.832				
Adjusted *R* Square	0.831				
Std. Error of the Estimate	0.370				
Observations	215				
	df	Sum of square	Mean square	*F*	Sig
ANOVA					
Regression	1	143.642	143.642	1050.653	$P<0.001$
Residual	213	29.121	0.136		
Total	214	172.763			

Table 3. The statistical test results of MLR model.

Regression statistics					
R	0.836				
R Square	0.699				
Adjusted *R* Square	0.697				
Std Error of the Estimate	0.494				
Observations	215				
	df	Sum of square	Mean square	*F*	Sig
ANOVA					
Regression	1	120.681	120.681	493.556	$P<0.001$
Residual	213	52.082	0.243		
Total	214	172.763			

For the training dataset, the R^2 and MSE of the MLR model was 0.70 and 0.25, respectively. However, the ANN model on the same training dataset produced much improved results with R^2=0.841 and MSE=0.133, respectively. For the testing dataset, the ANN model improved the R^2 of the MLR model by 12% and the MSE value by 30%. For the whole dataset, the MLR model produced a R^2 value of 0.699. This coefficient was improved to 0.832 with the ANN model. The MSE value of the MLR model was 0.243 which compares with a much-reduced value of 0.136 for the ANN model. Clearly, the ANN model can better predict skin permeability from Abraham descriptors (Table 4).

Dependence of skin permeability on the descriptors The ANN model was used to analyze the influence of each Abraham descriptor. Skin permeability was calculated using the ANN model by varying 1 Abraham descriptor each time while keeping the rest of the descriptors constant mean value of each range. The results are shown in Figure 3.

Discussion

Generally, log K_p has inversely depended on π_2^H, the solute polarity. The statum corneum (SC) lipid phase is mainly composed of hydrocarbon substance. It is well known that solute partition into the lipid phase decreases in the hydrocarbon solvents with increasing solute polarity. The relationship between skin permeability and the partition coefficient can be expressed by the following equation:

$$K_p = \frac{K_m \times D}{\delta} \quad (3)$$

where K_m is the skin-water partition coefficient of the solute, D is its diffusivity through the skin, and δ is the diffusion path length.

$\Sigma\alpha_2^H$ and $\Sigma\beta_2^H$ reflect solute hydrogen bonding activity. Many researchers have discussed the relationship between hydrogen bonding and skin permeability[15,19], and have suggested an inverse relationship between them. This was also observed in this study. The reason is similar as that of π_2^H; the increasing solute hydrogen bond acceptor and donor activity resulted in decreased partitioning into the organic phase due to the free energy cost associated with the disruption of hydrogen bonds[19]. It is interesting to note that in the MLR model there was a positive relationship between $\Sigma\alpha_2^H$ and log K_p. This suggests that although the MLR model has reasonable precision, it may not be necessary to provide correct relationships between skin permeability and some of the Abraham descriptors.

Table 4. Comparison between MLR and ANN models.

Model	Training set+Validation set			Testing set			Total dataset		
	n	R^2	MSE	*n*	R^2	MSE	*n*	R^2	MSE
MLR model	172	0.700	0.250	43	0.698	0.213	215	0.699	0.243
ANN model	172	0.841	0.133	43	0.792	0.149	215	0.832	0.136

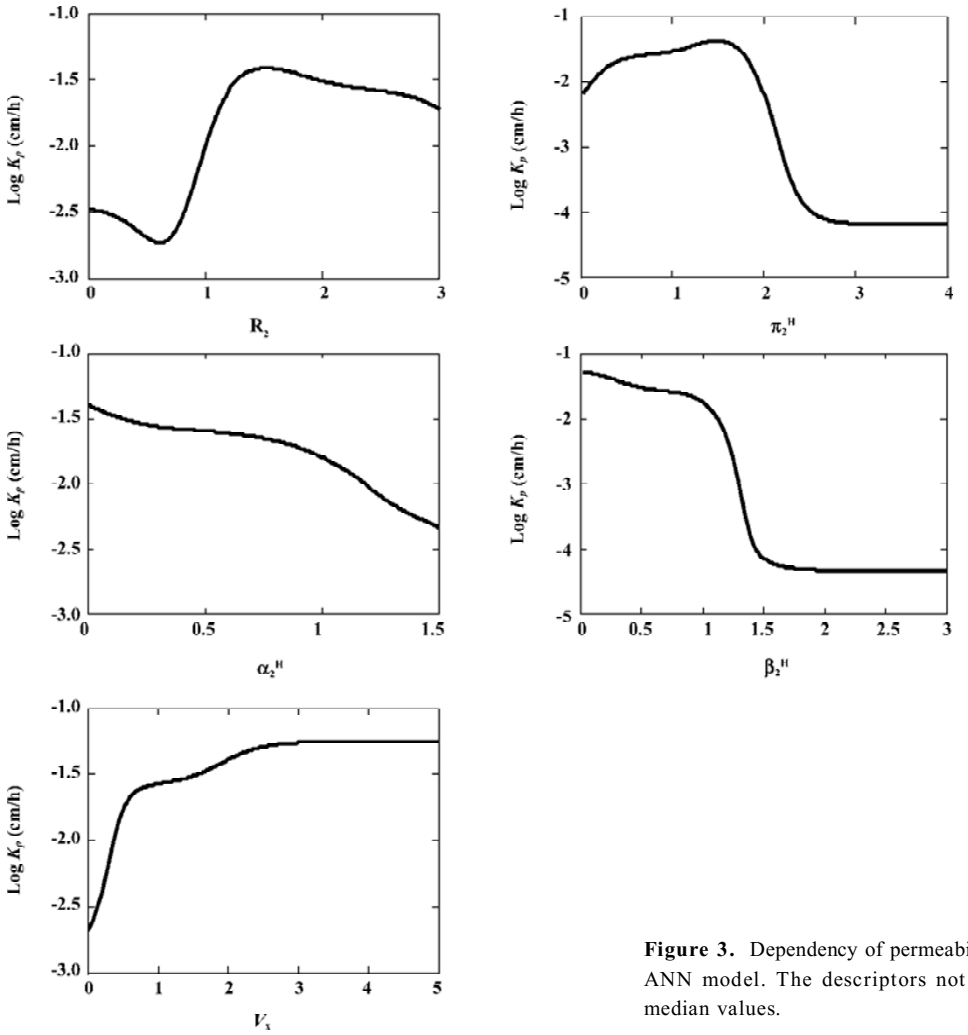

Figure 3. Dependency of permeability log K_p on the single descriptor for ANN model. The descriptors not shown in the plots were set to their median values.

Predicted skin permeability increased with the McGowan characteristic volume descriptor. This was not expected. A similar, unexpected relationship was observed before by Potts and Guy[19]. They explained that the molecular volume represented a combination of physical phenomena: the impact of molecular size on partition and diffusion. On the one hand, increasing molecular size increased the hydrophobic surface area and this increased the partitioning into a lipid phase. On the other hand, larger molecules diffused more slowly, leading to reduced permeability. The positive relationship between K_p and V_x tends to suggest that partitioning effects dominate.

The effect of R_2 on skin permeability is rather complex. In previous studies, Potts and Guy have suggested that the effect of R_2 was not significant[19].

An artificial neural network model for predicting skin permeability was developed and compared with a multiple linear regression model. The model was based on a comprehensive set of experimental data from various sources in the open literature. Both the MLR and ANN (ANN-5711) models are related to Abraham descriptors. Compared to the MLR model, the ANN model was shown to give better prediction results which indicate non-linearity and complexity of correlation between Abraham descriptors and skin permeability. Some insight into the degree of nonlinear behavior of every Abraham descriptors has also been assessed with a functional dependence to understand relationships.

References

1 Scheuplein RJ. Mechanism of percutaneous adsorption. I. Routes of penetration and the influence of solubility. J Invest Dermatol 1965; 45: 334–46.

2 Scheuplein RJ. Mechanism of percutaneous absorption. II. Transient diffusion and the relative importance of various routes of skin penetration. J Invest Dermatol 1967; 48: 79–88.

3 Scheuplein RJ, Blank IH. Permeability of the skin. Physiol Rev 1971; 51: 702–47.

4 Scheuplein RJ, Blank IH. Mechanism of percutaneous absorption. IV. penetration of nonelectrolytes (alcohols) from aqueous solutions and from pure liquids. J Invest Dermatol 1973; 60: 286–326.
5 CEFIC. CEFIC Workshop on methods to determine dermal permeation for human risk assessment. Utrecht: Institute of Occupational Medicine; 13–15 June 2004. Report No.: TM/04/07.
6 Guy RH, Hadgraft J. A theoretical description relating skin penetration to the thickness of the applied medicament. Int J Pharm 1980; 6: 321–32.
7 Wang TF. Microscopic models for the structure and permeability of the stratum corneum barrier layer of skin [PhD thesis], Buffalo: University at Buffalo, State University of New York, 2003.
8 Moss GP, Dearden JC, Patel H. Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. Toxicology *In vitro* 2002; 16: 299–317.
9 Martin YC. Quantitative drug design. New York: Marcel Dekker; 1978.
10 Ma XL, Chen C, Yang J. Predictive model of blood-brain barrier penetration of organic compounds. Acta Pharmacol Sin 2005; 26: 500–12.
11 Chen C, Yang J. MI-QSAR models for prediction of corneal permeability of organic compounds. Acta Pharmacol Sin 2006; 27: 193–204.
12 Potts RO, Guy RH. Predicting skin permeability. Pharm Res 1992; 9: 663–9.
13 Flynn GL. Physiochemical determinants of skin absorption. In: Gerity TR, Henry CJ, editors. Principles of route-to-route extrapolation for risk assessment. Amsterdam: Elsevier Science Publishing Co Inc; 1990. p 93–127.
14 Lien EJ, Gao H. QSAR analysis of skin permeability of various drugs in man as compared to *in vivo* and *in vitro* studies in rodents. Pharm Res 1995; 4: 583–7.
15 Pugh WJ, Degim T, Hadgraft J. Epidermal permeability-penetrant structure relationships: 4, QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. Int J Pharm 2000; 197: 203–11.
16 Degim T, Hadgraft J, Ilbasmis S, Ozkan Y. Prediction of skin penetration using artificial neural network (ANN) modeling. J Pharm Sci 2003; 92: 656–64.
17 Degim T. Understanding skin penetration: computer aided modelling and data interpretation. Current Computer-Aided Drug design 2005; 1: 11–9.
18 Abraham MH, Martins F, Mitchell RC. Algorithms for skin permeability using hydrogen bond descriptors: the problem of steroids. J Pharm Pharmacol 1997; 49: 858–65.
19 Potts RO, Guy RH. A predictive algorithm for skin permeability-the effects of molecular-size and hydrogen-bond activity. Pharm Res 1995; 12: 1628–33.
20 Johnson ME, 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: 1162–72.
21 Patel H, Berge WT, Cronin MTD. Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals, Chemosphere 2002; 48: 603–13.
22 PubChem Compound Browser. Bethesda (MD): National Library of Medicine (US); Available from URL: http://pubchem.ncbi.nlm.nih.gov.
23 Sontag ED. Feedback stabilization using two-hidden-layer nets. IEEE Transactions on Neural Networks 1992; 3: 981–90.
24 Boger Z, Guterman H. Knowledge extraction from artificial neural network models. IEEE Systems, Man, and Cybernetics Conference; 12–15 October 1997; Orlando, FL.