Correlation of blood-brain penetration and human serum albumin binding with theoretical descriptors

Mati Karelson, [*a,b,c] Dimitar Dobchev,[a,c] Tarmo Tamm,[c,d] Indrek Tulp,[b,c] Jaak Jänes,[b,c] Kaido Tämm,[b,c] Andre Lomaka,[a,c] Deniss Savchenko,[a,c] Gunnar Karelson[a,c]

[a] Department of Chemistry, Tallinn University of Technology, Akadeemia tee 15, Tallinn 19086, Estonia
[b] Institute of Chemistry, University of Tartu, Jakobi Street 2, Tartu 51014, Estonia
[c] MolCode, Ltd.; Soola 8, Tartu 51013, Estonia
[d] Institute of Technology, University of Tartu, Nooruse 1, Tartu 50411, Estonia
E-mail: mati.karelson@ttu.ee

Abstract

Quantitative Structure-Activity Relationship (QSAR) models were developed for blood-brain barrier and human serum albumin binding for a dataset of drugs where experimental values of both properties were available. All drugs were represented by chemical descriptors calculated from their constitutional, geometrical and topological structure, and quantum mechanical wave function. The obtained linear (multilinear regression) and nonlinear (artificial neural network) models link the drug structures to their reported properties. Also, based on the characterization of the descriptors we suggest additional criteria for the search of new active compounds. Each multilinear model was tested by leave-one-out and ABC methods. The latter method separates the all data points into three sets and predicts for each of them the property values. The former method is an iterative procedure which in each step it excludes one data point and predict its value based on the model rebuilt for the remaining data points. In addition, the predictive ability neural networks were assessed using the validation sets. All drug structures were investigated by conformational analysis in order to find the lowest energy conformers.

Keywords: QSAR, Neural Network, BMLR, Blood-brain barrier, Human serum albumin

---

Introduction

Precise pharmacokinetic, metabolic and toxicity information for a chemical compound is necessary for preclinical and clinical trials in drug discovery. To enhance the efficiency and cost-effectiveness of the pharmaceutical industry, in recent years, numerous *in vitro* ADME-Tox

assays[1-3] and *in silico* prediction methods[4-5] have been developed, improved and integrated into early-stage pharmaceutical research and development. Since the experimental methods for determing and evaluating ADME-Tox properties are often time-consuming and expensive, in silico predictive models are becoming increasingly important tools as relatively cheap alternatives for initial screening. The computational modeling of lead molecules enables screening and elucidation of the chemical structures and prediction of their ADME-Tox activities. In recent years, two- and three-dimensional methodologies for the delineation between molecular structure and property activity have been developed. A fully automated two-dimensional approach using the topological representation of the molecular structure is suitable for fast screening of virtual libraries.[6-8] Since conformational analyses and computationally intensive semi-empirical and *ab initio* calculations are considerably more time consuming, the three-dimensional optimization has particularly succeeded for molecular docking approaches.[9-10]

Human serum albumin

One of the main factors that influences the distribution, metabolism, pharmacological effect and elimination of a drug is its specific binding to human plasma proteins. The pharmacological activity of a drug, however, has been demonstrated to be related rather to unbound fraction of drug than total concentration.[11] The binding constants of various drugs to plasma proteins range over five log-units. Under various physiological and pathological conditions the amount of plasma proteins in blood might be decreased and for strongly bound drugs there could be risk of enhanced activity and undesirable side effects.[12-13]

Human serum albumin (HSA), a single monomeric polypeptide of 585 amino acidic residues, is the most abundant soluble protein in human serum, with a typical concentration of 0.6 mM in the bloodstream[14] contributing significantly to the colloidal oncotic pressure and antioxidative capacity of human serum.[15,16] Furthermore, HSA has a remarkable ability to bind, non-covalently, a wide range of chemically diverse endogenous (fatty acids, metabolites) and exogenous ligands.[17,18] HSA is composed of three homologous domains (I–III) each of which is divided into two subdomains (A and B).[19] The crystallographic studies of HSA-exogenous ligand complexes have shown that drug-like compounds will mainly interact with one of two high affinity binding sites on the protein sub domains IIA and IIIA, known also as the warfarin (Sudlow`s site I) and indole-benzodiazepine (Sudlow's site II) binding sites, respectively. They are similar in size and shape, possessing elongated hydrophobic pockets with cationic amino acid residues near their cavity entrances.[20]

During the past decade, numerous QSAR models have been developed to understand the HSA binding mechanisms and determine the structural features of drug-like compounds. The employment of hologram 2D QSAR methods based on several combinations of distinct molecular fragments produced a statitiscally excellent QSAR model (R^2=0.91, q^2=0.72) for the large training set (250 compounds) and test set (62 compounds) of structurally and therapeutically diverse compounds.[21] In addition, Gunturi et al.[22] using the stochastic optimisation method based on the ant colony systems and multiple linear regression, obtained a

QSAR model with good predictive power (R^2=0.80, q^2=0.75) for 94 compounds. The analysis of selected descriptors indicated the dependence of HSA affinity on hydrophobic interactions, solubility, size and shape of drug-like compound.[22] The preponderant contribution of hydrophobic regions of drugs to HSA binding is supported by QSAR models developed by Colmenarejo et al.[23] and Estrada et al.[24] for the same data set with significant correlation (R^2=0.82 and R^2=0.83, respectively). An electrotopological state QSAR model created by Hall et al.[25] performed satisfactorily in predicting HSA binding ability (R^2=0.77, q^2=0.70) for a set of 94 compounds demonstrating increased binding affinity due to the presence and electron accessibility of fluorine and chlorine atoms as well as aromatic, aliphatic methylene and hydroxy groups.[25] Moreover, Deeb et al.[26] and Votano et al.[27] employed a principal component-artificial neural network (PC-ANN) method using topological descriptors and achieved QSAR models with good correlation (R^2=0.85 for 56 compounds and R^2=0.90 for 808 compounds, respectively). A significant linear correlation (R^2=0.85, n=18) between plasma protein binding and hydrophobicity/lipophilicity (logD) for a series of acidic glycine antagonists 4-hydroxyquinolones was determined.[28] However, Kratochwil el al. using a structurally diverse set of molecules (n=76) demonstrated that the lipophilicity (logD) is rather poorly correlated (R=0.4 for acids and R=0.3 for bases) correlated to HSA binding.[29]

Blood brain barrier

The blood-brain barrier (BBB) is a complex membranous system of brain capillary endothelial cells, pericytes, astrocytes, and nerve endings that plays a central role in maintaining the homeostasis of the central nervous system by blocking the movement of all molecules except those relatively small molecules can cross cell membranes by means of lipid solubility (oxygen, carbon dioxide, ethanol, and steroid hormones) and those that are transported by specific transport systems (electrolytes, nucleosides, sugars and some amino acids). An additional cerebral protection from chemicals circulating in blood (endogenous toxins and various xenobiotics) is provided by the metabolic enzyme and active efflux transport systems (ATP binding cassette transporter super-family of proteins, organic anion transport proteins) in the membrane of the endothelial cells.[30]

In drug discovery and development, understanding of the mechanisms of penetration of drugs and drug-like compounds through the BBB is crucial to design new potential candidates to act on the central nervous system (CNS) and to avoid undesirable side effects of compounds acting in peripheral tissue.

In recent years, numerous *in vivo*, *in vitro* and *in silico* approaches have been applied to determine the penetration potential of drug candidates. For *in vivo* measurements imaging (MRI and PET),[31,32] autoradiography,[33] cerebrospinal fluid sampling[34] and microdialysis[35] techniques have been used. *In vitro* assessments have been carried out on isolated brain capillaries,[36] bovine brain microvessel endothelial cells,[37] Caco-2 cell monolayers,[38] MDR-MDCK cell lines,[39] parallel artificial membrane permeability assay (PAMPA)[40] and immobilized artificial membranes (IAMs).[41]

The blood-brain distribution is commonly expressed as the ratio of the steady state molar concentration of a compound in the brain and in the blood:

$BB = C_{brain}/C_{blood}$

During the last decade, a large number of computational models for the prediction of logBB were developed by several research groups (see Table 1).

Table 1. Summary of QSAR studies for the prediction of blood brain distribution

Descriptors used	No of compounds used in training set	R^2	s	References
R_2, π_2^H, $\alpha_2^H\beta_2^H$	57	0.91	0.197	[42]
E, S, A, B, VX, I1	148	0.745	0.343	[43]
HBAo, HBD	45, 70	0.72, 0.76		[44]
$N_{acc,Solv}$, log P, A_{pol}	61	0.730	0.424 (rms)	[45]
TPSA	45, 57	0.78, 0.66		[46]
SASA (Monte Carlo)	76	0.984	0.173	[47]
G_{Solv} (GB/SA)	55	0.722	0.37	[9]
E-state (3)	102	0.66	0.45	[6]
5 Structural descriptors (GA)*	59, 72	0.757,0.785	0.41, 0.358	[48]
MI-QSAR (5)	56	0.845		[49]
Structural descriptors (5)	48	0.837	0.26	[50]
log P, PSA, TIs	58	0.81		[51]
Atom type (3)	57	0.897	0.259	[52]
TOPS-MODE (2)	114, 81	0.697, 0.740	0.422	[7]
HAS (5)	47	0.781	0.375	[53]
V, Q_H, $Q_{O,N(MRA)}$	56	0.817	0.330	[53]
V, Q_H, $Q_{O,N(ANN)}$	56		0.236 (rmse)	[54]
C log P, TPSA	150	0.69		[55]
E-state index, AlogP98, van der Waals surface area, Kappa shape index	88	0.746	0.392	[56]

* Genetic algorithm

A less commonly used alternative to describe the brain permeation is the determination of the permeability-surface area coefficient (expressed as logPS), which reflects the free extracellular concentration of a chemical compound.[57] Recently, a QSAR model of logPS with significant

prediction ability ($R^2 = 0.87$, s = 0.52) for 30 neutral compounds was developed.[58] The comparison of mathematical equations for logPS and logBB revealed that the factors that determine logPS are also those that determine logBB.[59] Thus, the analysis of QSAR models indicates that lipophilicity and polar surface area are the most important molecular determinants of BBB permeation. Moreover, BBB permeation of passively diffusing drugs might be optimal when logP values are in the range of 1.5-2.7[60]. Additionally, for the successful design of a CNS drug, neutral or basic compounds (pKa=7.5-10.5) with relatively low molecular weight (<450) and small polar surface area (60-70Å^2), low number of hydrogen bond donors (<3) and acceptors (<7), as well as reduced molecular flexibility (number of rotatable bonds: <8) tend to be particularly suitable.[60]

The goal of the current study is to develop QSAR models for two biological properties based on a common chemical data set of drugs. The different models are interpreted in terms of descriptors involved and statistical parameters. Two approaches were taken in order to achieve this goal, namely, the development of multilinear mathematical equations and creation and training of artificial neural networks. Before QSAR development, all drugs were subjected to conformational search in order to find the optimum low energy structures.

Data Set

Blood-brain barrier partitioning (logBB)

The partitioning of a compound across the blood-brain barrier is experimentally measured as the ratio of the concentration of the compound in the brain, C_{brain}, to that in the blood, C_{blood}: BB = C_{brain} /C_{blood}. The data set investigated contains 60 compounds (mainly drugs) with their corresponding BB values, directly measured[6] that were converted to the logarithm scale for analysis (see Table2).

Binding Affinity to Human Serum Albumin (logHSA)

Serum protein binding is usually given as the percentage of a drug bound to plasma proteins at clinically achievable concentrations. These values are usually obtained using *in vitro* measurements at several concentrations, resulting in a calculated mean value. An alternative method, designed for more rapid screening, involves the use of immobilized HSA as the stationary phase in an HPLC (High-Performance Affinity Chromatography) procedure. In the present study, we use albumin binding affinity values assayed by HPLC by Colmenarejo *et al.*[23] were used. The dataset was optimized to maximize the diversity in both structure and physicochemical properties and it contains binding affinity constants *k (HSA)* for 85 diverse drugs collected in Table 2. The affinity binding constant *k* was computed from retention time (*t*) as follows: $k = (t - t_o)/t_o$, where t_o is the time for passage of a nonretained material.

Table 2. Experimental (obsd) and predicted (calcd) LogBB and logHSA of all drugs

No.	CAS number	Compound[b]	logBB obsd	log BB clacd BMLR	logBB calcd ANN	logHSA obsd	logHSA calcd BMLR	logHSA calcd ANN
1	37517-30-9	Acebutolol				-0.21	-0.036	-0.174
2	103-90-2	Acetaminophen	-0.31	-0.01	-0.093	-0.81	-0.756	-0.724
3	87848-99-5	Acrivastine				-0.02	0.323	0.596
4	18559-94-9	Albuterol	-1.03	-0.586	-0.602a			
5	28981-97-7	Alprazolam	0.04	0.389	0.244a			
6	13655-52-2	Alprenolol				0.04	0.266	0.055
7	58-15-1	Aminopyrine	0	-0.136	0.327			
8	50-48-6	Amitriptyline	0.89	0.926	1.02			
9	57-43-2	Amobarbital	0.04	0.067	-0.263			
10	26787-78-0	Amoxicillin				-1.21	-0.915	-0.751
11	60-80-0	Antipyrine	-0.1	0.091	-0.235	-0.69	-0.172	-0.576
12	50-78-2	Aspirin	-0.5	-0.884	-0.763a	-1.39	-1.061	-0.71a
13	29122-68-7	Atenolol	-0.87	-0.441	-0.444	-0.48	-0.433	-0.347
14	51-55-8	Atropine	-0.06	0.289	-0.138			
15	25614-03-3	Bromocriptine	-1.1	-0.94	-0.563			
16	10457-90-6	Bromperidol	1.38	1.134	1.075			
17	28395-03-1	Bumetanide				-0.03	0.054	-0.217
18	58-08-2	Caffeine	-0.06	0.092	-0.0505	-0.92	-0.881	-0.818
19	298-46-4	Carbamazepine	-0.14	-0.091	-0.107	-0.1	0.065	-0.128
20	36507-30-9	Carbamazepine epoxide	-0.35	-0.346	-0.399			
21	154-93-8	Carmustine	-0.52	-0.415	-0.223			
22	55268-75-2	Cefuroxime				-1.33	-1.253	-1.031
23	64544-07-6	Cefuroxime axetil				-0.56	-0.836	-0.85
24	15686-71-2	Cephalexin				-1.11	-0.48	-1.491
25	56-75-7	Chloramphenicol				-0.46	-0.859	-0.495a
26	50-53-3	Chlorpromazine	1.06	1.021	1.101	1.1	0.677	0.821
27	94-20-2	Chlorpropamide				-0.44	-0.295	-0.185
28	51481-61-9	Cimetidine	-1.42	-0.748	-1.142	-0.44	-0.32	-0.432
29	85721-33-1	Ciprofloxacin				0.14	-0.092	-0.086
30	22316-47-8	Clobazam	0.35	0.249	0.51			
31	637-07-0	Clofibrate				0.27	0.215	0.111
32	4205-90-7	Clonidine				-0.13	-0.153	0.116
33	23593-75-1	Clotrimazole				1.34	0.914	1.261
34	76-57-3	Codeine	0.55	0.165	0.100a			

Table 2. Continued

35	16110-51-3	Cromolyn				-1.07	-0.707	-0.836
36	50-47-5	Desipramine	1.2	0.962	0.728	0.61	0.565	0.648
37	439-14-5	Diazepam	0.52	0.46	0.478			
38	2095-95-6	Didesipramine	1.06	0.826	0.450a			
39	71-63-6	Digitoxin				0.13	0.392	0.479
40	548-73-2	Droperidol				0.43	0.678	0.508
41	50-28-2	Estradiol				0.68	0.159	0.233
42	33419-42-0	Etoposide				-0.49	-0.587	-0.389
43	2022-85-7	Flucytosine				-1.11	-0.729	-1.063
44	1622-62-4	Flunitrazepam	0.06	-0.237	-0.188			
45	69-23-8	Fluphenazine	1.51	1.312	1.127			
46	54-31-9	Furosemide				-0.13	-0.32	-0.333
47	10238-21-8	Glyburide				0.68	0.727	0.563
48	52-86-8	Haloperidol	1.34	1.062	1.052			
49	56-29-1	Hexobarbital	0.1	-0.068	-0.192			
50	58-93-5	Hydrochlorothiazide				-0.42	-0.456	-0.571
51	50-23-7	Hydrocortisone				-0.4	-0.086	-0.39
52	68-88-2	Hydroxyzine	0.39	0.595	0.243			
53	15687-27-1	Ibuprofen	-0.18	-0.737	-0.465a			
54	50-49-7	Imipramine	1.07	1.06	0.922	0.75	0.682	0.763
55	53-86-1	Indomethacin	-1.26	-0.322	-1.022	0.47	0.374	0.34
56	84625-61-6	Itraconazole				1.04	1.411	1.379
57	65277-42-1	Ketoconazole				0.84	0.806	0.954
58	22071-15-4	Ketoprofen				0.03	-0.034	-0.103
59	36894-69-6	Labetalol				0.14	0.019	0.002a
60	84057-84-1	Lamotrigine				-0.13	-0.173	-0.128
61	59-92-7	Levodopa	-0.77	-1.091	-0.749			
62	77-07-6	Levorphanol	0	0.179	0.139			
63	137-58-6	Lidocaine				-0.23	0.079	-0.101
64	91-84-9	Mepyramine	0.49	0.11	0.466			
65	5588-33-0	Mesoridazine	-0.36	-0.022	-0.485			
66	59-05-2	Methotrexate	-1.52	-1.887	-1.702a	-0.77	-0.61	-0.419
67	83-43-2	Methylprednisolone				-0.22	-0.082	-0.099
68	37350-58-6	Metoprolol				-0.29	-0.205	-0.08
69	10118-90-8	Minocycline				0.21	-0.272	0.238
70	57-27-2	Morphine	-0.16	-0.26	-0.228			
71	42200-33-9	Nadolol				-0.4	-0.434	-0.422
72	22204-53-1	Naproxen				0.25	-0.037	-0.161
73	22316-55-8	N-Desmethylclobazam	0.36	0.412	0.521a			

Table 2. Continued

	129618-40-							
74	2	Nevirapine	0	-0.212	-0.099			
75	1088-11-5	Nordazepam	0.5	0.503	0.473			
76	70458-96-7	Norfloxacin				0.14	-0.283	0.16
77	303-81-1	Novobiocin				0.35	0.167	0.105
78	82419-36-1	Ofloxacin				0.14	-0.126	-0.101
79	99614-02-5	Ondansetron				0.37	0.424	0.347
80	604-75-1	Oxazepam	0.61	-0.004	0.043			
81	6452-71-7	Oxprenolol				-0.15	0.063	-0.1
82	87-08-1	Penicillin V				-0.69	-0.568	-0.48
83	76-74-4	Pentobarbital	0.12	0.034	0.278			
84	77-10-1	Phencyclidine	0.68	0.92	1.33			
85	50-33-9	Phenylbutazone	-0.52	-0.128	-0.428	0.19	0.298	0.222
86	57-41-0	Phenytoin	-0.04	0.14	-0.075	0	-0.268	-0.121
87	57-47-6	Physostigmine	0.08	0.19	0.016			
88	13523-86-9	Pindolol				-0.13	-0.136	-0.071
89	19216-56-9	Prazosin				0.06	-0.182	-0.027
90	50-24-8	Prednisolone				-0.4	-0.092	-0.404
91	125-33-7	Primidone	-0.07	-0.46	-0.188a			
92	59-46-1	Procaine				-0.19	-0.019	-0.165
93	57-83-0	Progesterone				0.59	0.319	0.449a
94	58-40-2	Promazine	1.23	0.692	1.28	0.92	0.526	0.848
95	525-66-6	Propranolol	0.64	0.233	0.624	0.28	0.397	0.135
96	51-52-5	Propylthiouracil				-0.75	-0.675	-0.783
97	56-54-2	Quinidine	-0.46	-0.246	-0.228	0.44	0.569	0.452
98	130-95-0	Quinine				0.49	0.562	0.491
99	66357-35-5	Ranitidine				-0.1	-0.409	-0.321
100	69-72-7	Salicylic acid	-1.1	-0.669	-0.589	-0.66	-0.878	-0.858
101	487-54-7	Salicylurate	-0.44	-0.586	-0.503			
102	51-34-3	Scopolamine				-0.34	-0.247	-0.072a
103	3930-20-9	Sotalol				-0.44	-0.121	-0.339
104	599-79-1	Sulfasalazine				0.56	-0.009	0.28
105	14759-06-9	Sulforidazine	0.18	-0.218	0.339			
	103628-46-							
106	2	Sumatriptan				-0.05	-0.229	0.002a
107	321-64-2	Tacrine	-0.13	0.232	-0.308			
108	63590-64-7	Terazosin				-0.16	-0.285	-0.039a
109	91161-71-6	Terbinafine				1.17	1.248	0.947
110	58-22-0	Testosterone				0.74	0.276	0.296

Table 2. Continued

111	94-24-6	Tetracaine				0.32	0.174	0.097a
112	60-54-8	Tetracycline				-0.08	-0.526	-0.075
113	83-67-0	Theobromine	-0.28	-0.035	-0.703			
114	58-55-9	Theophylline	-0.29	-0.091	-0.438			
115	76-75-5	Thiopental	-0.14	-0.539	-0.081			
116	26839-75-8	Timolol				-0.33	-0.557	-0.411
117	1156-19-0	Tolazamide				-0.42	-0.473	-0.185
118	64-77-7	Tolbutamide				-0.22	-0.316	-0.353
119	28911-01-5	Triazolam	0.74	0.726	0.685			
120	117-89-5	Trifluoperazine	1.44	1.728	1.605a			
121	146-54-3	Triflupromazine				1.05	1.016	0.567a
122	738-70-5	Trimethoprim				-0.26	-0.245	-0.348
123	73-22-3	Tryptophan				-0.78	-0.659	-0.377a
124	52-53-9	Verapamil	-0.7	-0.638	-0.15	0.52	0.721	0.823
125	81-81-2	Warfarin				-0.04	0.557	-0.055
126	30516-87-1	Zidovudine				-1.02	-0.864	-1.175

[a] drugs used for validation set of the ANN, [b]-all smile strings for the compounds are available in supplementary material SM2

Methodology

Structure optimizations and descriptor calculations

In the current study, QSAR (Quantitative Structure-Activity Relationship) models are presented for log BB and log HSA involving theoretical descriptors, which were calculated solely from molecular (drug) structure using the QSARModel program.[61] These descriptors can be classified as: (i) constitutional, (ii) geometrical, (iii) topological, (iv) charge-related, (v) quantum chemical, and (vi) thermodynamic.[62] The total number of descriptors for each property ranged between 600 and 900 per compound.

A conformational search was also performed using the MacroModel software package.[63] For the calculations, MMFF94s - a static variant of Merck Molecular Force Field 94 (MMFF94) was used.[64] For energy minimization the Polak-Ribiere Conjugate Gradient (PRCG) method with a gradient 0.05 kcal/Å as a stopping criterion was used. Furthermore, for the conformational search, the Monte Carlo Multiple Minimum (MCMM) method was used where 100 steps per rotatable bond and up to 15000 steps per compound were defined.[65] Conformers with the minimum potential energy were utilized as the MOPAC 6 input structures.[66] Within quantum-mechanical, semi-empirical calculations, the AM1[67] parameterization was used with the gradient norm setting of 0.01 kcal/Å.

For reasons of comparison and completeness, optimization of the structures without prior conformational search was also performed. However, the models derived from these drug structures did not lead to significant QSAR equations and hence these results are not presented.

QSAR modeling

(a) Linear approach (BMLR). The Best Multilinear Regression method[68] (BMLR) was used to find the best correlation models from selected non-collinear descriptors. BMLR selects the best two-parameter regression equations, the best three-parameter regression equations, *etc*. on the basis of the highest R^2 and F values in the step-wise regression procedure. The result obtained by BMLR is the "best" linear representation of the property in the given descriptors pool.

(b)Nonlinear approach (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 associated set of output values.[69] Through an iterative "learning" process, the network refines the information derived from the input values (descriptors) in order to reproduce an associated set of property values (experimental logBB, logHSA). 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 this case, the signal emerging from the output neuron represents the current calculated log BB or log HSA value. When this value is compared to the desired (experimental target) property, a measure of the network error can be calculated. The root-mean squared error (RMS) was used to quantify the effectiveness of the network. 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 – logBB or structure-log HAS relationship.

A backpropagation network was used along with the delta rule for optimization of the weights.[9] In Figure 1 a typical structure of a fully connected neural network is shown with neurons that accept inputs I_i. All I_i form the total input for a neuron, Net, weighed by the connection weights w. Each neuron calculates a single output, Out, using Net transformed by a sigmoidal function. 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. In the training process the weights w are iteratively adjusted by an appropriate amount Δw, which is proportional to the network error with respect to each weight.
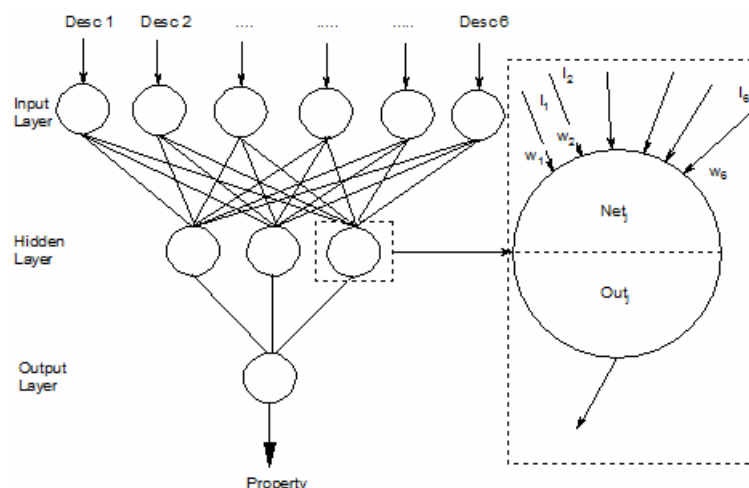
Figure 1. Fully connected three-layer neural network.

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 a 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.

In order to find the most important descriptors as inputs to the net, a sensitivity analysis was performed on a preselected descriptor space, based on the lowest (RMS). This space was formed after applying the following criteria for reduction of the total descriptor space: i) all descriptors with variance less than 10^{-4} were excluded, ii) descriptors which do not correlate with the property more than $r^2=0.2$ were excluded and iii) by chemical inspection of certain irrelevant descriptors (for instance, descriptors related to the internal entropy and enthalpy are less likely to be connected to the investigated property).

An important stage of the modeling process is to define the proper architecture of the ANN models. Several different architectures of ANN models were built for each property in this work. In the search for optimal ANN architecture the lowest possible number of neurons was sought, in order to follow the common principle of generality of the ANN prediction. In addition, the RMS for each different architecture was monitored (regarding the hidden units in the hidden layer) so that to select the one with the lowest RMS. The number of layers was chosen to be three-fold based on the basis of common practice for QSAR ANN modeling[70] and by taking into account the number of data points so as to reduce the possibility of overfitting during the training stage.

ABC validation of the MLR models

To validate the multilinear models, the data was sorted in ascending order according to the experimental values, three subsets (A, B, C) were then formed: the 1st, 4th, 7th, etc. data points comprise the first subset (A), the 2nd, 5th, 8th, etc. comprise the second subset (B), and the 3rd, 6th, 9th, etc. comprise the third subset (C). The three training sets were prepared as the combinations of any two subsets (A and B), (A and C), and (B and C), respectively. The tested MLR model was then rebuilt for each of the training sets with the same descriptors but optimized coefficients, and used to predict the property values of the respective C, B and A subsets. The prediction was assessed based on the R^2 between the predicted and experimental property values.[71]

In addition to the ABC validation, the standard leave-one-out (LOO) cross-validation (R^2_{cv}) for all developed models was used.

Results and Discussion

The BMLR algorithm was used to generate several multilinear equations for each property with the number of descriptors between 2 and 7. The final equations were selected by taking into account several factors in order to obtain significant models: (i) the number of descriptors in the models should follow the basic "rule of thumb" *i.e.* not less than five data points per descriptor and (ii) relevance of the descriptors toward the nature of the phenomenon under investigation and (iii) statistical feasibility of each additional descriptor (in terms of ΔR^2, ΔF and t-test value).

BMLR model for logBB

The MLR model obtained for logBB is presented as Eq.1 together with its statistical parameters and descriptors. This is the best equation found for the current data which had a significant quality of fit in terms of $R^2 = 0.81$.

$logBB = -(5.272\pm1.064)- (29.85\pm2.853)*D1+(0.341\pm0.0611)*D2- (9.676\pm1.505)*D3- (0.560 \pm0.105)*D4-(0.010\pm0.002)*D5+(0.268\pm0.066)*D6$ (1)

$N = 60$, $k = 6$, $R^2 = 0.810$, $R_{cv}^2 = 0.748$, $F = 37.69$, $s^2 = 0.112$;

The notations along Eq. 1 are as follows: N is the number of data points; k is the number of descriptors; R^2 is the squared correlation coefficient; R^2_{cv} is the squared cross-validated correlation coefficient (LOO); F is the Fisher's criterion; and s^2 is the squared standard error. Also, the standard errors of the regression coefficients are also indicated in the equation.

The experimental and predicted values of logBB together with their prediction bands at 95% confidence level (denoted as dashed lines) are shown in Figure 2. The numerical values of these data are shown in Table 2. As can be noted from Figure 2, there are two outliers according to these bands i.e. Methotrexate (underestimated) and Indomethacin (overestimated).
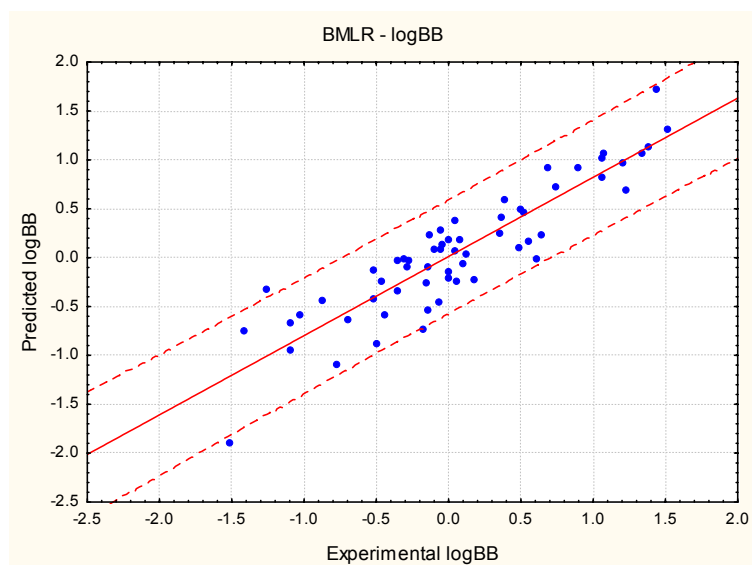
Figure 2. Experimental and predicted logBB values based on equation 1.

Equation (1) possesses six descriptors (D1-D6) that are listed in Table 3. According to the t-statistics of the descriptors, the most significant of them is D1 - *HA dependent HDCA-2 (AM1) (all)*. It is related to the hydrogen donor ability of the drug and the respective partial charge on the hydrogen surface area. The definition of this descriptor takes into account as many polar hydrogen atoms as there are possible hydrogen bond acceptors (N, O, S, P, F, Cl, Br and I atoms). Increasing the values of this variable leads to decreased logBB values. Alone, this descriptor correlates with the experimental logBB with $R^2 = 0.37$. Its maximum and minimum values are 0.0984 and 0.00291, respectively. Hydrogen bonding was reported[11] to be one of the most important factors for the blood-brain barrier penetration.

Table 3. Descriptors involved in BMLR models (1) and (2)

Desc ID	Descriptor name[a]
D1	HA dependent HDCA-2 (AM1) (all)
D2	Number of halogenated groups
D3	Maximum net atomic charge (Zefirov) for N atoms
D4	HOMO-1 energy (AM1)
D5	WFOSA Atomic charge (AM1) weighted FOSA
D6	Number of amino groups
D7	Number of C atoms
D8	Partial Surface Area of O atoms
D9	Charged (Zefirov) Surface Area of H atoms
D10	Difference (Pos - Neg) in Charged Part of Partial Charged Surface Area (AM1)
D11	Minimum net atomic charge (Zefirov) for any atom type

[a]-descriptor values are available in supplementary material SM1

Further, the next pair of descriptors D3 and D6 reflects the influence of the presence of N atoms in the drug molecules on logBB. Almost all compounds (except 3 compounds) of the current data set for the log BB include N atoms; hence the related descriptors appear in the QSAR model. The maximum net atomic charge for N atoms (D3) is the second statistically most significant descriptor according to the descriptor t-value. Its value range is within the interval [0.00915, -0.125]. As can be seen from this range, the sign of this descriptor value changes depending on drugs compound. The number of amino groups D6 varies from 0 to 3.

The next group of descriptors D2, D4 and D5 (with numerical ranges [0,3], [-11.46, -8.77], [0, 100.6]) is connected to the short–range and hydrophobic interactions of the drug molecules while penetrating the blood brain barrier. It is known that halogenated compounds are more lipophilic and poorly soluble in water, which leads to favorable penetration of the drug through the barrier. This fact is reflected by the positive sign of the coefficient of D2 in Eq. 1. D4 describes the availability/reactivity of electron pairs (as H-bond acceptors); D5 is a measure of the polarity of the hydrophobic part of the solvent accessible surface area, as it is weighted by the partial charges of the hydrogen and sp^3 hybridized carbon atoms.

In addition to the "rule of five", we recommend when searching (screening) for drugs with high logBB to combine the following characteristics based on the above descriptors: i) hydrogen-donor ability D1<0.002 ii) D2=2, two halogenated benzene groups (preferably with attached F) connected together with a nonrotatable bridge, iii) D6 >3, three tertiary amino groups located spatially as close as possible and iv) maximum net charge on N atoms, D3 < -0.1.

Regarding the domain of the applicability of Eq. 1, it can be defined based on the ranges of the independent and dependent variables. Since the data set used to build Eq. 1 is diverse, we suggest that for reliable prediction of logBB for novel drugs candidates, the descriptor values of these should be within the range $[Di_{min} - f.|Di_{min}|, Di_{max} + f.|Di_{max}|]$, where the "applicability" coefficient f = 0.3 and i=1-6. The value of the latter coefficient was determined empirically by taking into account the overall predictive ability of Eq. 1 for the current number of datapoints and its physico-chemical reasoning. Note that the domain of applicability for (1) is not the same as statistical applicability defined by the standard deviation at certain confidence level. It involves a wider chemical space composed from the current drug set plus the external drugs set. Furthermore, the predictions of logBB that lead to deviations from the interval $[logBB^{obsd}_{min} - f.|logBB^{obsd}_{max} - logBB^{obsd}_{min}|, logBB^{obsd}_{max} + f.|logBB^{obsd}_{max} - logBB^{obsd}_{min}|]$ should be carefully analyzed and if necessary discarded. It is worth mentioning that the above minimum and maximum values for Di and logBB(observed) concern the current data set.

In principle applicability domain of a model is defined only by the descriptors in the equations. However, because of the significant diversity of the BB set we have also included LogBB (obsd) limits in order to be sure for the correctness of external predictions. In our case, the whole applicability domain consists of a cube in 7-dimensional space spanned over six descriptors plus the property. To give a sense of the applicability (domain) bounds, Figure 3 presents the scatter plot between the predicted by Eq. 1 LogBB (open blue circles) and the descriptor values of D1. In Figure 3, it is shown the two-dimensional rectangular which defines

the upper and lower bounds according to the limits ([-2.429, 2.419; 0.00204, 0.128]) described in the above paragraph with coefficient of applicability 0.3. As can be seen from Figure 3, all predicted LogBB are within the 2-dimensional rectangular of the domain of applicability. However, for external predictions, points outside this rectangular must be analyzed.
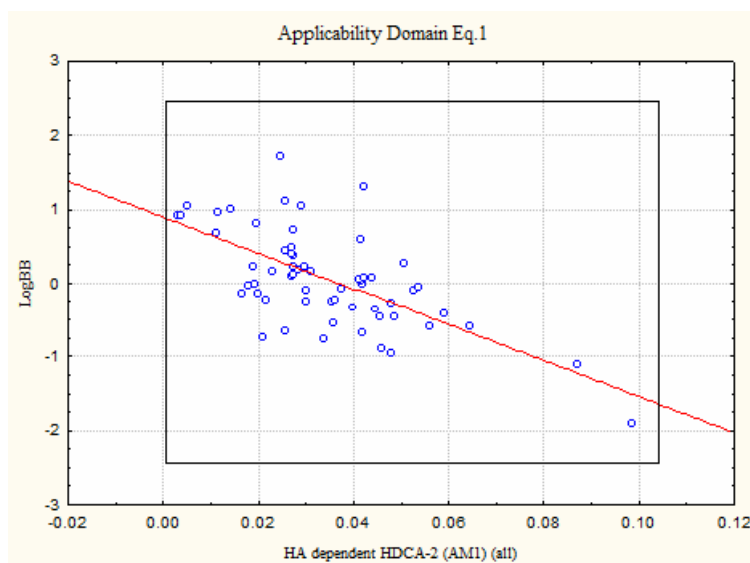

Figure 3. Applicability domain of equation 1 according to D1 descriptor and the augmented LogBB(obsd) bounds.

BMLR model for logHSA

The second BMLR model concerned a data set of 85 drugs with experimental logHSA values. The best MLR model obtained for logHSA was a 5-descriptor equation as expressed by Eq. 2. The experimental and predicted logHSA values for this model are collected in Table 2. The graphical presentation of the correlation between these values is shown in Figure 4 together with the prediction bands at the 95% confidence level. The statistical criteria for Eq. 2 indicate good correlation bearing in mind the diversity of the drugs as well as the number of data points.

logHSA = − (0.515±0.173) + (0.097±0.008)*D7 − (6.093±0.586)*D8 − (0.117±0.015)*D9 − (9.016±1.355)*D10 − (6.349±1.402)*D11 (2)
N = 85, k = 5, R^2 = 0.812, R^2_{cv} = 0.785, F = 68.06, s^2 = 0.072
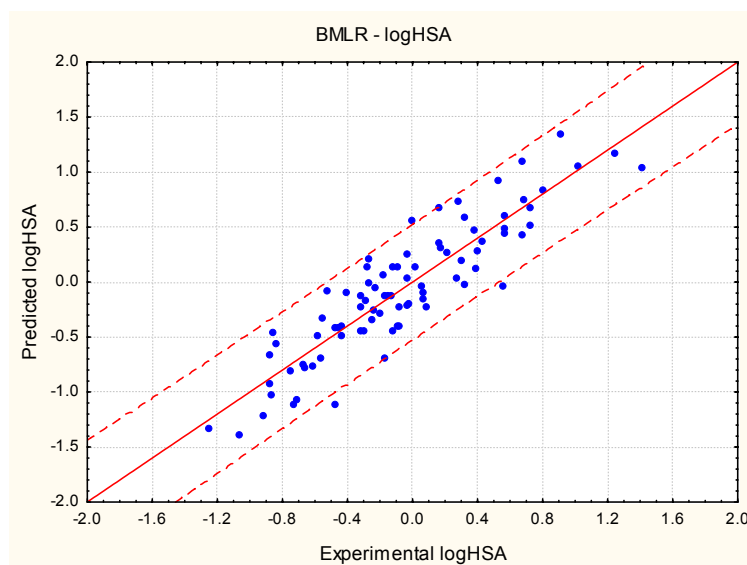
Figure 4. Experimental and predicted logHSA values based on Eq. 2.

The two statistically most significant descriptors according to t-values in Eq. 2 are D9 and D10. These two molecular features are related to the charged surface area and the total difference between the charged positive and negative charged surface areas suggesting their importance for the HSA binding phenomenon. In addition, D9 ($|t_{D9}|>|t_{D10}|$) is related to the hydrogen atoms which indicates that the hydrogen-bonding capacity of the drugs plays an important role for in HSA binding. In this regard, it is worth mentioning that D8 – *Partial surface area of O atoms* appears also in Eq.2. The remaining two descriptors D11 and D7 characterize the molecule as a whole. However, the latter descriptor is related to the total weight of the compound and hydrophobicity. According to its sign, it suggests that the logHSA would increase with the increase of D7 for the current set used.

Similarly to Eq.1, the domain of applicability can be defined for Eq. 2. The descriptor ranges for D7-D11 are [4, 41], [0, 0.352], [5.582, 24.48], [-0.195, 0.065] and [-0.188, -0.099], respectively. The same analyses for Eq. 1 hold for Eq. 2. However, Eq. 2 is based on a larger number of data points and hence the applicability coefficient for the prediction of novel compounds can be slightly extended *i.e.* f = 0.4.

Since the research involving HSA goes in both directions *i.e.* seeking for drugs with highly binding affinity and compounds with low binding affinity, we therefore could choose to search for active compounds with high logHSA satisfying the following simultaneous constrains of the descriptors: i) D9>13, ii) compounds with more separated positive minus negative charged surface *i.e.* D10>0.02, iii) number of C atoms D7>22, iv) compounds with lower O surface, D8<0.1.

It is important to note that for both models the constraints defined for search of active compounds must be within the limits of the minimum and maximum descriptor values defined by the applicability coefficient f. For example, for Eq. 2 and D7 is said to be searched for active

compounds with D7>22 that implies the search to be carried out within the range of D7 [22, $D7_{max}$ + f.$D7_{max}$]. In addition, it should be taken into account the optimum (for large logHSA, logBB) between the groups of descriptors, which lead to positive and negative contributions in equations 1 and 2.

ABC validation for MLR models of logBB and logHSA
One of the main goals of this study was also to develop as general as possible MLR models in order to cover a larger chemical space of applicability. Therefore, for the sake of generality, we did not use the division of data into training and external test sets. Instead of direct external validation, a type of leave-many-out validation, *i.e.* the ABC validation (see Methodology) was employed to "mimic" external validation. The efficiency of QSAR models to predict the scales of log BB and logHSA was assessed by the squared correlation coefficients $R^2_{(Pred)}$ between experimental and predicted data for test sets (A, B or C). The results from the ABC validation are presented in Table 4. The overall assessment of the predictions for both models (1) and (2) is significant as can be noted by the closeness between the average R^2, $R^2_{(Fit)}$ and $R^2_{(Pred)}$ values.

Table 4. Statistical characteristics for ABC validation of the multilinear models 1 and 2.

TRAINING SET	N	$R^2_{(Fit)}$	$R^2cv_{(Fit)}$	TEST SET	N	$R^2_{(Pred)}$
BMLR model for logBB						
A+B	40	0.826	0.734	C	20	0.766
A+C	40	0.836	0.749	B	20	0.737
B+C	40	0.806	0.694	A	20	0.798
Average		0.823	0.726			0.767
BMLR model for logHSA						
A+B	57	0.808	0.765	C	28	0.802
A+C	57	0.829	0.789	A	28	0.761
B+C	56	0.810	0.769	B	29	0.803
Average		0.816	0.774			0.788

N – number of compounds used in the validation models
$R^2_{(Fit)}$ – squared correlation coefficient of the multilinear equation for the binary sets used as training sets (A+B, A+C or B+C)
$R^2_{cv(Fit)}$ – cross-validated squared correlation coefficient of the multilinear equations for the binary sets (leave-one-out approach)
$R^2_{(Pred)}$ – squared correlation coefficient for the regression between the predicted values (from the models of the binary sets) for the test sets and the respective experimental values

The prediction ability of Eq. 1 according to leave-one-out cross-validated R^2_{cv} (LOO) = 0.748 is significantly high. However, a significant difference (ΔR^2 = 0.061) between R^2_{cv} to R^2

may suggest that Eq. 1 is sensitive toward the exclusion of certain drug(s) during the LOO procedure. A possible reason is that the set is quite diverse and thus the exclusion of certain data points influences the prediction by the model. However, the quality of the average predictions increases when the model is examined using the more robust ABC validation (see Table 4).

Neural network model for logBB

Before the neural network treatment was started, the experimental logarithmic values and descriptor values were both normalized to a range 0-0.9 for internal consistency. Then the significant descriptors were selected by reducing the initial descriptor pool as described in the Methodology part. For each property, the available experimental data were divided into training and test sets. To preserve the generality of the network models, all test sets consisted of not more than ten data points. Further, sensitivity analyses were performed on the reduced descriptor space by constructing 1-1-1 neural networks and then the descriptors that produced the lowest RMS error were selected. Several neural network models with different architecture were investigated for each property.

The best model found for logBB had an architecture of 4-3-1. The input descriptors of this model are: *Number of halogen groups*, *SQRT (HA dependent HDCA-2)*, *Maximum net atomic charge for N atoms* and *Maximum electrophilic reactivity index (AM1) for H atoms*. As can be noted from Table 3, there is, quite similarity between the nature of the descriptors for the BMLR and ANN models. Hence, similar physico-chemical analysis holds for the ANN descriptors. Also, the *HA dependent HDCA-2* descriptor appeared as square root function indicating the nonlinear relation with the property.

The statistical parameters in terms of R^2 for this model resulted in $R^2_{train} = 0.85$ and $R^2_{val} = 0.87$ (RMS 0.26 and 0.32, respectively). The graphical presentation of these correlations between the experimental and predicted for both training and validation sets are shown in Figure 5. The numerical values for this ANN model are included in Table 2.
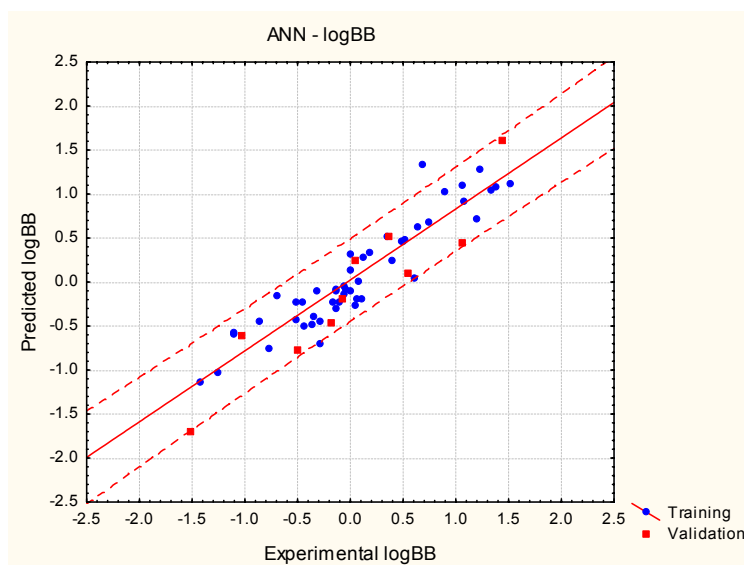
Figure 5. Experimental and predicted logBB values based on 4-3-1 ANN model (Training set 50 and validation set 10 data points).

Neural network model for logHSA

The next ANN model developed for the logHSA had the architecture 4-3-1 while the input neurons consisted of the following descriptors: *HASA-2/SQRT(TMSA) (AM1), Number of C atoms, Number of O atoms, Lowest e-e repulsion (1-center) (AM1) for Cl atoms*. This model was trained up to 411 iterations and the results from the predictions for both, training and test sets are shown in Figure 6 and Table 2. The statistical parameters in terms of R^2 for this model resulted in $R^2_{train} = 0.88$ and $R^2_{val} = 0.922$ and the achieved RMS error 0.20 and 0.32 for the training and validation sets, respectively.
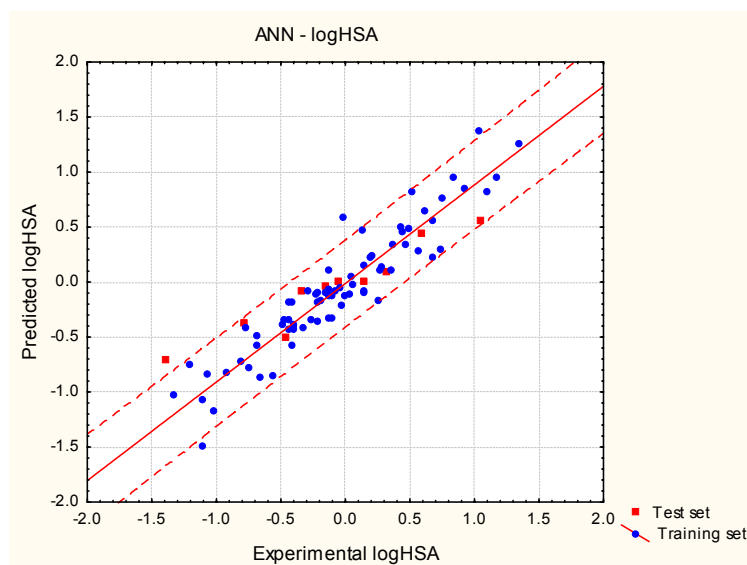
Figure 6. Experimental and predicted logHSA values based on 4-3-1 ANN model (Training set 75 and validation set 10 data points).

As can be noted most descriptors for this ANN model have similar chemical content as in the case of the linear model (C, O and Cl atoms). Also, *HASA-2/SQRT(TMSA)* is related to hydrogen-bonding acceptor charged surface area. However, the incorporation of nonlinear transformation by ANN gives a significant improvement of the QSAR model.

Generally, the applicability domains of both ANN models for logHSA and logBB are difficult to define compared to the BMLR models. However, because of the similar nature of the descriptors the same limits for the BMLR models can be also applied for these ANN models with the respective applicability coefficients.

As was shown in Figure 3, similar domain presentation can be done for both ANN models. However, because of the better predictivity of the ANN models their coefficient of applicability can be extended to 0.4-0.5.

Conclusions

The overlapping chemical data set of 126 drugs allows development of QSAR correlations for two biological properties *i.e.* blood-brain barrier and human serum albumin binding with their theoretical descriptors. The results obtained indicate that both the multilinear regression and ANN models exhibit reasonable prediction capabilities. While the linear models were developed mainly for the purpose of structure-activity interpretation, the ANN models were primarily developed for predictions and classification. It is worth noting that the drug structures obtained by conformational search resulting in energy minimized conformers significantly improved both the BMLR and ANN models.

The applicability domains were defined for both BMLR and ANN models for searching of active compounds with predetermined property values, Therefore, the models reported should enable prediction and screening for analogous drugs for their blood-brain barrier penetration and human serum albumin binding.

Supplementary Information Available

SM1 − Descriptor values available for all models. SM2 − smile strings available for all compounds in Table 2.

References and Notes

1. Stenberg, P.; Norinder, U.; Luthman, K.; Artursson, P. *J. Med. Chem.* **2001**, *44*, 1927.
2. Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. *J. Pharmacol. Sci.* **2001**, *90*, 749.
3. Cucullo, L.; Hossain, M.; Rapp, E.; Manders, T.; Marchi, N.; Janigro, D. *Epilepsia* **2007**, *48*, 505.
4. Gunturi, S. B.; Narayanan, R. *QSAR Comb. Sci.* **2007**, *26*, 653.
5. Hall, L. M.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2120.
6. Rose, K.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 651.
7. Cabrera, M. A.; Bermejo, M.; Perez, M.; Ramos, R. *J. Pharm. Sci.* **2004**, *93*(*7*), 1701.
8. Konovalov, D. A.; Coomans, D.; Deconinc, E.; Heyden, Y. V. *J. Chem. Inf. Model.* **2007**, *47*, 1648.
9. Keserü G. M.; Molnar L. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 120.
10. Dessalew, N.; Patel, D. S.; Bharatam, P. V. *J. Mol. Graph. Model.* **2007**, *25*, 885.
11. Benet, L.Z.; Kroetz D. L.; Sheiner L. B. *Pharmacokinetics. The dynamics of drug absorption, distribution and elimination.* Goodman GA Ed., New York: McGraw-Hill, 1996, pp 3-27.
12. Steele, W. H.; Lawrence, J. R.; Stuart, J. F. B.; McNeill, C. A. *Cancer Chemother. Pharmac.* **1981**, *7(1)*, 61.
13. Kishino, S.; Nomura, A.; Itoh, S.; Nakagawa, T.; Takekuma, Y.; Sugawara, M.; Furukawa, H.; Todo, S.; Miyazaki, K. *Eur. J. Clin. Pharmacol.* **2002**, *58*, 621.
14. Kragh-Hansen, U. *Pharmacol. Rev.* **1981**, *33*, 17.
15. Evans, T. W. *Aliment Pharmacol. Ther.* **2002**, *16*, 6.
16. Armstrong, J. S.; Rajasekaran, M.; Hellstrom, W. J.; Sikka, S. C. *J. Andrology* **1998**, *19*, 412.
17. Tawara, S.; Matsumoto, S.; Matsumoto, Y. Kamimura, T.; Goto, S. *J. Antibiot.* **1992**, *45*, 1346.

18. Paal, K.; Müller, J.; Hegedus, L *Eur. J. Biochem.* **2001**, *268*, 2187.

19. He, X. M.; Carter, D. C. *Nature* **1993**, *358*, 209.

20. Petitpas, I.; Bhattacharya, A. A.; Twine, S.; East, M.; Stephen, C. *J. Biol. Chem.* **2001**, *276*, 22804.

21. Moda, T. L.; Montanari, C. A.; Andricopulo, A. D. *Letters in Drug Design & Discovery* **2007**, *4*, 502.

22. Gunturi, S. B.; Narayanan, R.; Khandelwal, A. *Bioorg. Med. Chem.* **2006**, *14*, 4118.

23. Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J.-L. *J. Med. Chem.* **2001**, *44*, 4370.

24. Estrada, E.; Uriarte, E.; Molina, E.; Simon-Manso, Y.; Milne, G. W. A. *J. Chem. Inf. Model.* **2006**, *46*, 2709.

25. Hall, L. M.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2120.

26. Deeb, O. B.; Hemmateenejad, B. *Chem. Biol. Drug. Des.* **2007**, *70*, 19.

27. Votano, J. R.; Parham, M.; Hall, L. M.; Hall, L. H.; Kier, L. B.; Oloff, S.; Tropsha, A. *J. Med. Chem.* **2006**, *49*, 7169.

28. Rowley, M.; Kulagowski, J. J.; Watt, A. P.; Rathbone, D.; Stevenson, G. I.; Carling, R. W.; Baker, R.; Marshall, G. R.; Kemp, J. A.; Foster, A. C.; Grimwood, S.; Hargreaves, R.; Hurley, C.; Saywell, K. L.; Tricklebank, M. D.; Leeson, P. D. *J. Med. Chem.* **1997**, *40,* 4053.

29. Kratochwil, N. A.; Huber, W.; Müller, F.; Kansy, M.; Gerber, P. R. *Biochem. Pharmacol.* **2002**, *64*, 1355.

30. Adenot, M. A. *Practical Approach to Computational Models of the Blood–Brain Barrier. In: Handbook of Neurochemistry and Molecular Neurobiology*, Springer US 2001, pp 109-150.

31. Lüdemann, L.; Hamm, B.; Zimmer, C. *Magn. Reson. Imaging* **2000**, *18*, 1201.

32. Luurtsema, G.; Molthoff, C. F. M.; Schuit, R. C.; Windhorst, A. D.; Lammertsma, A. A.; Franssen, E. J. F. *Nucl. Med. Biol.* **2005**, *32*, 87.

33. Ross, T. M.; Martinez, P. M.; Renner, J. C.; Thorne, R. G.; Hanson, L. R.; Frey II. *J. Neuroimmunol.* **2004**, *151*, 66.

34. Shen, D. D.; Artru, A. A.; Adkison, K. K. *Adv. Drug Deliv. Rev.* **2004**, *56*, 1825.

35. Mano, Y.; Higuchi, S.; Kamimura, H. *Biopharm. Drug Dispos.* **2002**, *23*, 351.

36. Siakotos, A. N.; Rouser, G.; Fleischer, S. *Lipids* **1969**, *4*, 234.

37. Otis, K. W.; Avery, M. L.; Broward-Partin, S. M.; Hansen, D. K.; Behlow, H. W. Jr.; Scott, D. O.; Thompson, T. N. *J. Pharmacol. Toxicol. Meth.* **2001**, *45*, 71.

38. Faassen, F.; Vogel, G.; Spanings, H.; Vromans, H. *Int. J. Pharm.* **2003**, *263*, 113.

39. Wang, Q.; Rager, J. D.; Weinstein, K.; Kardos, P. S.; Dobson, G. L.; Li, J.; Hidalgo, I. J. *Int. J. Pharm.* **2005**, *288*, 349.

40. Di, L.; Kerns, E. H.; Fan, K.; McConnell, O. J.; Carter, G. T. *Eur. J. Med. Chem.* **2003**, *38*, 223.

41. Yoon, C. H.; Kim, S. J.; Shin, B. S.; Lee, K. C.; Yoo, S. D. *J. Biomol. Screen* **2006**, *11*, 13.

42. Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1257.

43. Platts, J. A.; Abraham, M. H.; Zhao, Y. H.; Hersey, A.; Ijaz, L.; Butina, D. *Eur. J. Med. Chem.* **2001**, *36*, 719.

44. Oesterberg, T.; Norinder, U. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1408.

45. Feher, M.; Sourial, E.; Schmidt, J. M. *Int. J. Pharm.* **2000**, *201*, 239.

46. Ertl, P.; Rohde, B.; Selzer, P. *J. Med. Chem.* **2000**, *43*, 3714.

47. Kaznessis, Y. N.; Snow, M. E.; Blankley, C. J. *J. Comput. Aided Mol. Des.* **2001**, *15*, 697.

48. Hou, T.; Xu, X. *J. Mol. Model.* **2002**, *8*, 337.

49. Iyer, M.; Mishra, R.; Han, Y.; Hopfinger, A. *J. Pharm. Res.* **2002**, *19*, 1611.

50. Adenot, M.; Lahana, R. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 239.

51. Subramanian, G.; Kitchen, D. B. *J. Comput. Aided Mol. Des.* **2003**, *17*, 643.

52. Sun, H. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 748.

53. Stanton, D. T.; Mattioni, B. E.; Knittel, J. J.; Jurs, P. C *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1010.

54. Fu, X. C.; Wang, G. P.; Liang, W. Q.; Yu, Q. S. *Pharamazie* **2004**, *59*, 126.

55. Pan, D.; Iyer, M.; Jianzhong, L.; Li, Y.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2083.

56. Narayanan, R.; Gunturi, S. B. *Bioorg. Med. Chem.* **2005**, *13*, 3017.

57. Martin, I. *Drug Discovery Today* **2004**, *9*, 161.

58. Abraham, M. H. *Eur. J. Med. Chem.* **2004**, *39*, 235.

59. Goodwin, J. T.; Clark, D. E. *J. Pharm. Exp. Therap.* **2005**, *315*, 477.

60. Pajouhesh, H.; Lenz, G. R. *NeuroRx* **2005**, 2, 541.

61. www.molcode.com

62. (a)Karelson, M. *Molecular Descriptors in QSAR/QSPR*. Wiley & Sons: New York, 2000. (b) Balaban, A.; Pompe, M. *J. Phys. Chem.* **2007**, *111*, 2448. (c) Oliferenko, A.; Pisarev, S.; Palyulin, V.; Zefirov, N. *Adv. Quant. Chem.* **2006**, *51*, 139.

63. MacroModel, version 9.5, Schrödinger, LLC, New York, NY, 2007.

64. Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720.

65. MacroModel, version 9.5 User Manual, Schrödinger, LLC, New York, NY, 2007.

66. Stewart, J. J. P. Mopac 6.0, *QCPE* program *455*. *QCPE* **1990**, *9*, 10.

67. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

68. Katritzky, A. R.; Lobanov, V.; Mu, L.; Karelson, M. *J. Phys. Chem.* **1996**, *100*, 10400.

69. Haykin, S. *Neural Networks. A comprehensive foundation*. Pearson, India; 1999, pp 156-256.

70. Halberstam, N.; Baskin, I.; Paljulin, V.; Zefirov, N. *Russ. Chem. Rev.* **2003**, *72*, 629.

71. Katritzky, A. R.; Dobchev, D. A.; Fara, D. C.; Hur, E.; Tämm, K.; Kurunczi, L.; Karelson, M.; Varnek, A.; Solov'ev, V. P. *J. Med. Chem.* **2006**, *49*, 3305.