JPP Journal of Pharmacy and Pharmacology
ROYAL PHARMACEUTICAL SOCIETY

JPP Journal of Pharmacy and Pharmacology

JPP 2011, 63: 1411–1427
© 2011 The Authors
JPP © 2011 Royal Pharmaceutical Society
Received April 6, 2011
Accepted July 20, 2011
DOI
10.1111/j.2042-7158.2011.01345.x
ISSN 0022-3573

Research Paper

The application and limitations of mathematical modelling in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes

Gary P. Moss[a], Yi Sun[b], Simon C. Wilkinson[c], Neil Davey[b], Rod Adams[b], Gary P. Martin[d], M. Prapopopolou[d] and Marc B. Brown[e]

[a]School of Pharmacy, Keele University, Keele, [b]School of Engineering & Information Science, University of Hertfordshire, Hatfield, [c]Medical Toxicology Centre, Wolfson Unit, Medical School, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne, [d]Pharmacy Department, King's College London, Stamford Street, London and [e]MedPharm Ltd, Surrey Research Park, Guildford, UK

Abstract

Objectives Predicting the rate of percutaneous absorption of a drug is an important issue with the increasing use of the skin as a means of moderating and controlling drug delivery. One key feature of this problem domain is that human skin permeability (as K_p) has been shown to be inherently non-linear when mathematically related to the physicochemical parameters of penetrants. As such, the aims of this study were to apply and evaluate Gaussian process (GP) regression methods to datasets for membranes other than human skin, and to explore how the nature of the dataset may influence its analysis.

Methods Permeability data for absorption across rodent and pig skin, and artificial membranes (polydimethylsiloxane, PDMS, i.e. Silastic) membranes was collected from the literature. Two quantitative structure–permeability relationship (QSPR) models were used to compare with the GP models. Further performance metrics were computed in terms of all predictions, and a range of covariance functions were examined: the squared exponential (*SE*), neural network (*NNone*) and rational quadratic (*QR*) covariance functions, along with two simple cases of Matern covariance function (*Matern3* and *Matern5*) where the polynomial order is set to 1 and 2, respectively. As measures of performance, the correlation coefficient (CORR), negative log estimated predictive density (NLL, or negative log loss) and mean squared error (MSE) were employed.

Key findings The results demonstrated that GP models with different covariance functions outperform QSPR models for human, pig and rodent datasets. For the artificial membranes, GPs perform better in one instance, and give similar results in other experiments (where different covariance parameters produce similar results). In some cases, the GP predictions for some of the artificial membrane dataset are poorly correlated, suggesting that the physicochemical parameters employed in this study might not be appropriate for developing models that represent this membrane.

Conclusions While the results of this study indicate that permeation across rodent (mouse and rat) and pig skin is, in a statistical sense, similar, and that the artificial membranes are poor replacements of human or animal skin, the overriding issue raised in this study is the nature of the dataset and how it can influence the results, and subsequent interpretation, of any model produced for particular membranes. The size of the datasets, in both absolute and comparative senses, appears to influence model quality. Ideally, to generate viable cross-comparisons the datasets for different mammalian membranes should, wherever possible, exhibit as much commonality as possible.

Keywords Gaussian processes; in-vitro methods; percutaneous absorption; quantitative structure–permeability relationships

Correspondence: Gary P. Moss, School of Pharmacy, Keele University, Keele, Staffordshire ST5 5BG, UK.
E-mail: g.p.j.moss@mema.keele.ac.uk

Introduction

In-vitro drug penetration studies normally involve the use of excised human skin, skin from suitable animals or synthetic model barrier membranes. Studies using excised human skin provide a good indication of drug penetration, especially through the *stratum*

corneum, the primary barrier to drug penetration. Porcine skin has been widely and effectively employed as a substitute to human skin by many researchers.[1,2] Various synthetic membranes, and those of biological origin, have also been employed in drug release studies, notably dialysis membranes[3] and Silastic® (polydimethylsiloxane (PDMS)).[4–13] Such membranes must be capable of simulating in-vivo conditions if they are to be used to examine steady-state drug diffusion kinetics.

PDMS has been widely utilised as a model barrier membrane for determining drug diffusion from transdermal devices. Nacht and Yeung[6] demonstrated its lipid-like properties and compliance with diffusion kinetics in accordance with Fick's first law, which make it an acceptable model membrane. However, it has also been suggested that PDMS may be limited in representing skin absorption, particularly across a wide range of physicochemical properties.[13] In addition, it has been shown that, when quantitative structure–permeability relationship (QSPR) models are derived for PDMS membranes they are mechanistically different from those derived for human skin.[14,15]

Nevertheless, a major limitation to in-vitro diffusion studies through excised human skin is the availability of suitable tissue. This has resulted in the widespread use of synthetic membranes and various animal models in place of excised human skin, and may impact upon the ability to develop representative models based on in-vitro experimental data.[16] Studies comparing drug penetration across the skin of various animals with drug penetration across human skin have indicated that the skin of several animals provide excellent models for human skin. The models best mimicking drug release through human skin are pig and monkey skin.[1,17] Further, Garrett and Chemburkar[4] demonstrated that drug release through PDMS provided a good comparison with excised skin. Despite its advantages, excised human or animal skin cannot duplicate certain in-vivo effects, including metabolism, blood flow and removal of the penetrant from the site of application. When Cronin *et al*.[14] modelled permeation across a PDMS membrane they determined that flux of the penetrant was related to its ability to form hydrogen bonds, and not to its lipophilicity, as suggested by a number of studies on skin *ex vivo*.[15,18] Similar findings were later reported by Geinoz *et al*.[19]

Human skin differs from that of many animals in the thickness of its *stratum corneum*, the number of appendages per unit area and the amount of lipids present in the skin. Despite this, it is very unusual that no quantitative mathematical models have been developed for the purpose of characterising permeation across non-human skin. This is perhaps due to the development and success of the Potts and Guy model in 1992,[18] the first major model for predicting percutaneous absorption which was based on human skin data. However, the widespread use of animal skin in experiments since 1992, which is often validated by comparison with human skin data, only provides partial validation as it does not specifically examine the mechanistic nature of the absorption process, as quantitative models can. Further, despite extensive work on comparing the permeation across different species, the use of rodent skin in particular continues to be found extensively throughout the literature.

Analysis of human skin permeability has been shown to be non-linear when mathematically related to the physicochemical parameters of penetrants.[16,20,21] These studies have also shown that non-linear methods, such as Gaussian processes (GPs), have statistically outperformed other methods, such as QPSRs. Further, they have shown that GP models provided more accurate predictions of permeability than QSPR models.[16,21] Therefore, the aims of this study were twofold; firstly, to apply and evaluate GP methods, by comparison with QPSR models, to datasets for membranes other than human skin, which are commonly used as reasonable replacements. Secondly, this work aims to highlight potential pitfalls in the analysis of such data, and to show how the simple application of models may provide outcomes and findings that are not always as straightforward as they may seem.

Methods

Description of the datasets employed

The four datasets employed in this study have been collated from a number of literature sources. The human, pig, rodent and artificial membrane datasets (the latter being comprised of PDMS membranes and cultured membranes, as specified in Table 1) consist of 140, 15, 103 and 19 chemical compounds, respectively. While the human data has been presented extensively in the literature, and predominately consist of the well known and frequently studied Flynn dataset,[22] the others are shown in Table 1. Among these four datasets, there exist chemicals common to each dataset. Table 2 shows the number of chemical compounds common to human and animal or synthetic membranes.

Normally, most QSPR models suggest that log P and molecular weight (MW) are the most significant descriptors for permeability. However, Roberts and others[82–87] have shown that other parameters, particularly hydrogen bonding, are of importance. This has also been shown recently.[16,20,88] Indeed, Lam *et al*.[16] suggested that certain parameters were effectively interchangeable, or possibly co-linear, and that removing or adding certain physicochemical descriptors to the model did not affect its statistical quality or its ability to accurately predict permeability. Hence, the current study employs those methods used previously[16,20,88] and examines five physicochemical descriptors – log P, MW, the number of hydrogen bond donor and acceptor groups and solubility parameter, the latter being defined by Fedors.[89] Detailed descriptions of the methods employed to determine these parameters may be found elsewhere.[13,16,20,21]

Visualisation of the skin datasets

To visualize the data, all four datasets were collated and then normalised to ensure that all five physicochemical descriptors employed had a zero mean and unit variance. Principal component analysis (PCA) was used to visualize the normalised data by mapping it onto a low-dimensional space with a linear transformation, where permeability (log K_p) was plotted against the first two principal components (Figure 1). PCA indicates that the first principal component is responsible for 43.0% of the total variance, and the second principal component for 31.4% (PCA1 = 0.34MW + 0.15logP + 0.36SP +

Table 1 Datasets for pig, rodent and polydimethylsiloxane (PDMS) membranes, collated from the literature. Key physicochemical parameters used for model development are also shown, where relevant or available, for each chemical

Name	MW	MPt (°C)	log P/logP_{KOWWIN}	Solubility parameter $(calcm^{-3})^{[0.5]}$	HA	HD	Membrane type	K_p (cm/h)	logK_p
2-Phenylphenol[23]	170.21	59	3.28	12.24	1	1	Pig skin	0.0159	−1.80
Atropine[24]	289.38	116	1.91	11.04	3	1	Pig skin	0.00000176	−5.75
Captopril[13]	217.29	106	0.84	11.55	2	2	Pig skin	0.00147	−2.83
Captopril butyl ester[13]	273.29	127	4.38	8.14	2	0	Pig skin	0.0108	−1.97
Captopril ethyl ester[13]	245.29	105	3.39	9.25	2	0	Pig skin	0.0081	−2.09
Captopril hexyl ester[13]	301.29	150	5.36	9.07	2	0	Pig skin	0.00122	−2.91
Captopril methyl ester[13]	231.29	119	2.90	9.31	2	0	Pig skin	0.00359	−2.44
Captopril pentyl ester[13]	287.29	138	4.87	9.11	2	0	Pig skin	0.00108	−2.97
Captopril propyl ester[13]	259.00	116	3.89	9.20	2	0	Pig skin	0.0122	−1.91
Ligustrazine hydrochloride[25]	136.10	113.48	2.18	10.24	2	0	Pig skin	0.0002363	−3.63
Mannitol[26]	182.17	138.97	−3.01	18.63	6	6	Pig skin	0.00152	−2.82
Salicylic acid[27]	138.10	158.00	2.24	14.39	2	2	Pig skin	0.00127	−2.90
Hyoscine (scopolamine)[24]	303.40	59.00	0.39	11.53	4	1	Pig skin	0.00000051	−6.29
Tropicamide[24]	284.40	96.50	1.19	12.53	3	1	Pig skin	0.0000007	−6.15
Water[26]	18.02	0.00	−1.38	26.68	1	1	Pig skin	0.00225	−2.65
17a-Hydroxyprogesterone[28]	330.50	276	3.08	10.98	3	1	Rodent skin	0.00087	−3.06
2-(2-Methoxyethoxy) ethanol[29]	120.15	−70	−1.18	10.25	3	1	Rodent skin	0.08	−1.10
2,4 Dimethylamine[30]	266.13	86	0.84	9.52	3	2	Rodent skin	0.00031	−3.51
2-Ethoxyethanol[31]	90.12	−90	−0.42	10.33	2	1	Rodent skin	0.000148	−3.83
4-Methylaniline[32]	107.20	43.7	1.62	10.58	1	1	Rodent skin	0.08549	−1.07
4-n-Butylaniline[32]	149.00	−14	3.10	9.51	1	1	Rodent skin	0.23	−0.64
Alachlor[33]	269.77	40	3.37	9.80	2	0	Rodent skin	0.000655	−3.18
Alizapride[34]	339.90	207	1.80	12.06	6	2	Rodent skin	0.0057	−2.24
Aminopyrene[35]	231.00	108	0.60	10.7	3	0	Rodent skin	0.033	−1.48
Aniline[32]	93.00	−6.2	1.08	10.83	1	1	Rodent skin	0.067	−1.17
Atrazine[33]	215.69	175	2.82	11.77	5	2	Rodent skin	0.000769	−3.11
Benzyl acetate[36]	150.18	−51.3	2.08	10.10	1	0	Rodent skin	0.00027	−3.57
Bisphenol A diglycidyl ether[37]	340.80	10	3.84	10.38	4	0	Rodent skin	5.50E-06	−5.26
Bufexamac[35]	223.30	154	1.98	12.43	3	2	Rodent skin	0.271	−0.57
Butyl salicylate[38]	194.23	−6	4.08	11.45	2	1	Rodent skin	0.00002	−4.7
Caffeine[39]	194.20	238	0.16	32.83	4	0	Rodent skin	0.001025	−2.99
Clebopride[34]	373.90	162	3.21	11.47	4	2	Rodent skin	0.0074	−2.13
Clotrimazole[27]	344.85	148	6.26	11.17	2	0	Rodent skin	0.0055	−2.26
Coumarin[40]	146.15	70.6	1.51	11.91	1	0	Rodent skin	0.00079	−3.1
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT)[41]	354.49	108.50	6.79	9.45	0	0	Rodent skin	0.0068	−2.17
Decabromodiphenyl oxide (DBDPO)[42]	959.17	302.5	12.11	8.10	1	0	Rodent skin	0.00000703	−5.15
Decane[29]	142.30	−29.7	5.25	7.58	0	0	Rodent skin	9.30E-05	−4.03
Di(2-ethylhexyl)phthalate (DEHP)[43]	390.57	−50	8.39	9.39	2	0	Rodent skin	0.000946	−3.02
Diethylphthalate (DEP)[44]	222.24	−3	2.65	10.51	2	0	Rodent skin	0.00037	−3.43
Dibutyl squarate[45]	226.27	176.71	2.45	10.58	4	0	Rodent skin	0.00085	−3.07
Dibutylphthalate[44]	278.35	−35	5.11	10.86	2	0	Rodent skin	0.000089	−4.05
Diethyl squarate[45]	170.16	131.55	4.07	11.50	4	0	Rodent skin	0.001	−3.00
2-Sec-butyl-4,6-dinotrophenol (Dinoseb)[46]	240.22	40	3.67	11.10	2	1	Rodent skin	0.00115	−2.94
dodecane[29]	170.34	−9.6	6.23	7.69	0	0	Rodent skin	0.000014	−4.85

Table 1 *(Continued)*

Name	MW	MPt (°C)	log P/logP_{KOWWIN}	Solubility parameter $(calcm^{-3})^{[0.5]}$	HA	HD	Membrane type	K_p (cm/h)	logK_p
Dodecyl glycidyl ether (C12GE)[29]	242.20	58	5.01	8.41	2	0	Rodent skin	0.000029	−4.54
Domperidone[34]	425.92	242.5	3.35	12.48	3	2	Rodent skin	0.0028	−2.55
Ethanol[47]	46.07	−114.1	−0.14	10.92	1	1	Rodent skin	0.000415	−3.38
Ethyl benzene[29]	106.20	−94.9	3.03	9.04	0	0	Rodent skin	0.00031	−3.51
Etorphine[48]	411.55	215	3.02	11.76	5	2	Rodent skin	0.0046	−2.34
Felodipine[49]	384.30	283	3.86	10.43	3	1	Rodent skin	0.004	−2.40
Fenoxapropethyl[50]	361.78	84	4.95	11.06	5	0	Rodent skin	0.00071	−3.15
Hydroquinone[51]	110.11	170	1.03	15.18	2	2	Rodent skin	2.26E-05	−4.66
Ketoprofen[35]	254.29	94	3.12	11.75	2	1	Rodent skin	0.0187	−1.73
Lidocaine[52]	234.34	67	1.66	8.78	2	1	Rodent skin	0.0089	−2.05
Linoleic acid[53]	289.45	−5	7.51	9.05	1	1	Rodent skin	0.000029	−4.54
Mannitol[54]	182.17	138.97	−3.01	18.53	6	6	Rodent skin	0.00023	−3.64
Mefenamic acid[35]	241.29	230	5.28	11.85	2	0	Rodent skin	0.0077	−2.11
Methanol[55]	32.04	−98	−0.63	11.68	1	1	Rodent skin	0.0018	−2.74
Methyl nicotinate[56]	137.14	42.5	0.64	11.80	2	0	Rodent skin	0.00222	−2.65
Metopimazine[34]	445.61	170.5	2.42	13.55	3	1	Rodent skin	0.0051	−2.29
Morphine[57]	285.30	255	0.72	13.68	4	2	Rodent skin	0.00015	−3.82
Naphthalene[29]	128.20	80.6	3.17	10.42	0	0	Rodent skin	0.00051	−3.29
n-Butanol[55]	74.14	−89.8	0.84	10.13	1	1	Rodent skin	0.0115	−1.94
n-Heptanol[55]	116.20	−34	1.82	9.57	1	1	Rodent skin	0.1015	−0.99
n-Hexanol[55]	102.18	−52	2.03	9.71	1	1	Rodent skin	0.0374	−1.43
Nicardipine[49]	479.54	136	3.90	10.87	5	1	Rodent skin	0.0049	−2.31
Nicorandil[58]	211.18	92.5	0.43	14.40	3	1	Rodent skin	0.000727	−3.14
Nifedipine[49]	346.30	172	4.04	10.92	4	0	Rodent skin	0.0017	−2.77
Nimodipine[49]	418.40	125	3.13	10.50	3	1	Rodent skin	0.0026	−2.59
Nitrendipine[49]	360.40	184.13	2.99	11.44	5	1	Rodent skin	0.0039	−2.41
N-N-Diethyl m-toluamide[59]	191.28	−45	2.26	10.70	1	0	Rodent skin	0.00168	−2.77
n-Octanol[55]	130.23	−15.5	2.81	9.45	1	1	Rodent skin	0.097	−1.01
Nonane[29]	128.30	−53.5	4.76	7.51	0	0	Rodent skin	0.000042	−4.38
o-Cresyl glycidyl ether (oCGE)[37]	164.20	30.25	2.16	10.38	2	0	Rodent skin	0.000134	−3.87
Paraquat[47]	257.16	300	−2.71	10.45	0	0	Rodent skin	0.000346	−3.46
Prednisolone 21-decanoate[60]	514.70	145	6.07	11.67	5	2	Rodent skin	0.000111	−3.95
Prednisolone 21-heptanoate[60]	472.62	186	4.60	12.02	5	2	Rodent skin	0.000077	−4.11
Prednisolone 21-nonanoate[60]	500.68	131	5.58	11.78	5	2	Rodent skin	0.000092	−4.04
Prednisolone 21-octanoate[60]	486.65	159	5.09	11.89	5	2	Rodent skin	0.00011	−3.96
Prednisolone 21-pentadecanoate[60]	584.84	139	6.79	11.23	5	2	Rodent skin	0.000098	−4.01
Prednisolone 21-tridecanoate[60]	556.78	140	6.02	11.39	5	2	Rodent skin	0.000106	−3.97
Prednisolone 21-undecanoate[60]	528.73	129	5.54	11.57	5	2	Rodent skin	0.000149	−3.83
Progesterone[28]	314.50	121	3.67	10.05	2	0	Rodent skin	0.011	−1.96
Propranolol[61]	295.00	233.2	0.74	11.96	3	2	Rodent skin	0.00005	−4.30
Salicylamide[38]	137.14	140	1.03	16.60	2	2	Rodent skin	0.000017	−4.77
Salicylic acid[62]	138.10	158	2.24	14.39	2	2	Rodent skin	0.01046	−1.98
Hyoscine (scopolamine)[34]	303.40	59	0.39	11.53	4	1	Rodent skin	0.0041	−2.39
Squaric acid[45]	114.06	293	−0.44	20.47	4	2	Rodent skin	0.0007	−3.15

Sucrose[63]	342.30	190	−4.27	18.06	11	8	Rodent skin	0.0022	−2.66
TDCPP[66]	430.91	27	3.65	8.67	1	0	Rodent skin	0.00056	−3.25
Theophylline[64]	180.17	273	−0.39	14.05	4	1	Rodent skin	0.0022	−2.66
Thiourea[65]	76.12	176	−1.31	15.23	0	2	Rodent skin	9.60E-05	−4.02
Toluene (methyl benzene)[29]	92.10	−94.9	2.54	9.14	0	0	Rodent skin	0.0011	−2.96
Triclosan[66]	289.55	56	2.47	10.02	2	1	Rodent skin	1.3609	0.13
Tridecane[29]	185.40	−5.5	6.73	7.74	0	0	Rodent skin	0.000015	−4.82
Trifluralin[33]	335.28	49	5.31	9.49	6	0	Rodent skin	0.000279	−3.55
Undecane[29]	156.31	−25.6	5.74	7.64	0	0	Rodent skin	0.000025	−4.60
Urea[67]	60.60	135	−1.56	14.36	1	2	Rodent skin	0.000016	−4.80
Water[26]	18.02	0	−1.38	26.68	1	1	Rodent skin	0.00198	−2.7
Xylene (dimethyl benzene)[29]	106.20	−50	3.09	9.10	0	0	Rodent skin	0.00017	−3.77
Propoxur[68]	209.25	87	1.90	10.31	2	1	Rodent skin	0.00375	−2.43
Testosterone[69]	288.40	155	3.27	10.66	2	1	Rodent skin	0.0018	−2.74
1,6-Hexanediol diglycidyl ether (HDDGE)[37]	230.20	84.83	0.84	9.36	4	0	Rodent skin	0.000402	−3.40
2-Phenoxyethanol[70]	138.17	14	1.10	11.49	2	1	Rodent skin	0.001764	−2.75
2-Phenylphenol[23]	170.21	59	3.28	12.24	1	1	Rodent skin	0.0266	−1.58
5-Fluorouracil[71]	130.01	281	−0.81	13.46	3	2	Rodent skin	0.0000603	−4.22
Atenolol[72]	266.30	147	−0.03	12.49	4	3	Rodent skin	0.0301	−1.52
Benzoic acid[43]	122.10	122	1.87	11.94	1	1	Rodent skin	0.00952	−2.02
Bromopride[34]	344.26	152.5	1.94	10.74	4	2	Rodent skin	0.0078	−2.11
Corticosterone[28]	346.50	183	1.99	11.91	4	2	Rodent skin	0.00053	−3.28
Cortisone[73]	360.50	220	1.81	12.10	5	2	Rodent skin	0.00047	−3.33
Deoxycortisone[28]	330.47	141.5	3.12	11.03	3	1	Rodent skin	0.0034	−2.47
Diethylene glycol monomethyl ether[29]	120.20	−14.6	−1.50	10.25	3	1	Rodent skin	0.08	−1.09
Ethylaniline[32]	121.20	−64	2.11	9.73	1	1	Rodent skin	0.115	−0.94
Haloperidol[74]	375.90	151.5	4.22	10.78	4	1	Rodent skin	0.02	−1.70
Hydrocortisone[75]	362.50	220	1.62	12.75	5	3	Rodent skin	0.0001	−4.00
Lorazepam[76]	321.16	167	2.41	12.91	3	2	Rodent skin	0.0005662	−3.25
Metoclopramide[34]	354.30	182	1.69	11.13	4	2	Rodent skin	0.0091	−2.04
Amethocaine (tetracaine)[77]	264.37	152	3.02	9.85	3	1	PDMS(SIlastic®)	0.00000149	−5.82
Benzocaine[77]	165.19	89	1.8	11.11	2	1	PDMS(SIlastic®)	0.00000135	−5.87
Benzoic acid[78]	122.10	122.4	1.87	11.94	1	1	Test Skin	0.086	−1.07
Caffeine[78]	194.20	238	0.16	32.83	4	0	Test Skin	0.02	−1.70
Captopril[13]	217.29	106	0.84	11.55	2	2	PDMS(SIlastic®)	0.00394	−2.4
Captopril butyl ester[13]	273.29	127	4.38	8.14	2	0	PDMS(SIlastic®)	0.0411	−1.386
Captopril ethyl ester[13]	245.29	105	3.39	9.25	2	0	PDMS(SIlastic®)	0.00986	−2
Captopril hexyl ester[13]	301.29	150	5.36	9.07	2	0	PDMS(SIlastic®)	0.107	−0.91
Captopril methyl ester[13]	231.29	119	2.9	9.31	2	0	PDMS(SIlastic®)	0.00488	−2.31
Captopril pentyl ester[13]	287.29	138	4.87	9.11	2	0	PDMS(SIlastic®)	0.0675	−1
Captopril propyl ester[13]	259.00	116	3.89	9.20	2	0	PDMS(SIlastic®)	0.0222	−1.65
Clotrimazole[27]	344.85	148	6.26	11.17	2	0	Graft Skin	1.88	0.27
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT)[79]	354.50	108.5	6.79	9.45	0	0	Test Skin	0.007	−2.15
Dibucaine[77]	343.47	418	4.04	10.7	4	1	PDMS(SIlastic®)	0.000000631	−6.19
Flufenamic acid[80]	281.20	133	4.88	10.96	5	2	Epidermal Equivalent	0.003053	−2.51
Ketocaine[77]	291.43	132	3.6	9.38	2	0	PDMS(SIlastic®)	0.00000034	−6.47
Procaine[77]	236.31	61	1.99	10.34	3	1	PDMS(SIlastic®)	0.000000114	−6.94
Salicylic acid[27]	138.10	158	2.24	14.39	2	2	Graft Skin	4.55	0.66
Testosterone[81]	288.40	155	3.27	10.66	2	1	PDMS(SIlastic®)	0.029	−1.54

HA and HD, number of hydrogen bond acceptor and donor groups respectively; K_p, permeability coefficient; log P/log P_{KOWWIN}, octanol-water partition coefficient; MPt, melting point; MW, molecular weight.

Table 2 The number of chemical compounds common to human and animal (synthetic) membranes

Dataset	Number of common members	Number of non-common members
Pig	3	12
Rodent	48	55
Synthetic (polydimethylsiloxane, PDMS)	7	12

0.63HA + 0.57HD; PCA2 = –0.56MW – 0.64logP + 0.47SP – 0.03HA + 0.23HD). Figure 1 also shows that there is no linear relationship between log K_p and the first or second principal components, suggesting the presence of more complex underlying non-linear patterns within the data. Further, it can be seen that the dataset for rodent skin has a similar distribution to human skin. Such comparisons are difficult to make for the pig and artificial membrane data as most of the log K_p values are outside the same range.

Quantitative structure–permeability relationship (QSPR) models

Two QSPR models were applied to the human skin dataset to provide a comparison with GP methods. The first model was that proposed by Potts and Guy,[18] where log K_p (cm/s) = 0.71logP – 0.0061MW – 6.3. The second model was that proposed by Moss and Cronin,[90] where log K_p (cm/h) = 0.74logP – 0.0091MW – 2.39. The latter model is derived from a slightly larger dataset than the former, and one where the original steroid data used by Potts and Guy,[18] as collated by Flynn,[22] has been corrected by Johnson.[91] It should also be noted that the above models used different methods for calculating log P. The absence of QSPR models for animal skin (apart from, for example, the work of Lin *et al*.[92] who examined the permeation of eight amino acids and dipeptides through porcine skin) meant that comparison in these cases were made against the naïve model.[20]

Gaussian process regression

A GP regression is defined as a collection of random variables that, jointly, have a Gaussian distribution and which is characterised completely by its mean and covariance function. The mean function is usually defined as the 'zero everywhere' function. The covariance function, $k(x_i, x_j)$, allows for specifying a-priori knowledge from a training dataset. It defines nearness or similarity between the values of $f(x)$ at the two points x_i and x_j.[20,93]

To make a prediction, y*, at a new input, x*, the conditional distribution, $p(y^*|y_1 \ldots, y_N)$, is computed on the observed vector $[y_1 \ldots, y_N]$. Since the model being applied is a GP, this distribution is also Gaussian in nature and is therefore completely defined by its mean and covariance. By the application of linear algebra, the mean and variance, x*, are given by

$$E[y_*] = \mathrm{k}_*^T\left(\mathrm{K} + \sigma_n^2\mathrm{I}\right)^{-1}\mathrm{y},$$
$$var[y_*] = k(\mathrm{x}_*, \mathrm{x}_*) - \mathrm{k}_*^T\left(\mathrm{K} + \sigma_n^2\mathrm{I}\right)^{-1}\mathrm{k}_*, \quad (1)$$

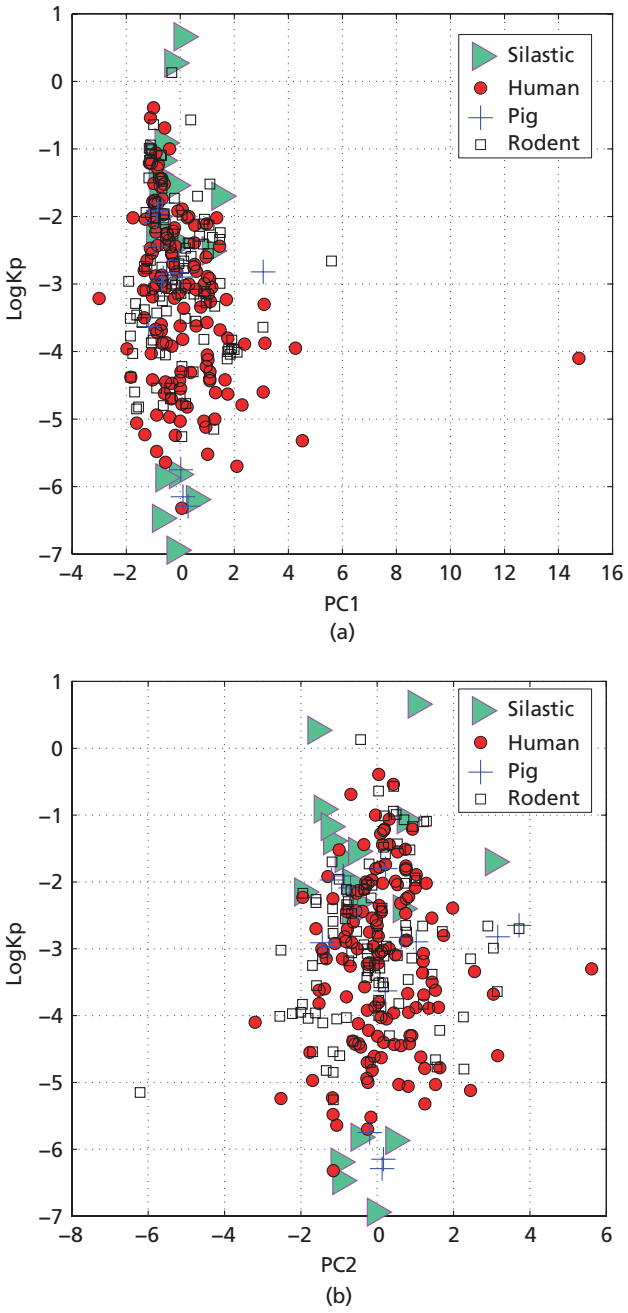

Figure 1 The relationship between log K_p and the principal component analysis (PCA) space of chemical compounds. (a) The first principal component. (b) The second principal component.

Where k is the covariance matrix; I the identity matrix; k* denotes the vector of covariances between the test point and the *N* training data; σ_n^2 denotes the variance of an independent identically distributed Gaussian noise; y denotes the vector of training targets; and $k(x^*, x^*)$ denotes the variance of y*. The mean is used as the predicted value generated by the GP model and the variance is shown as error bars on a graphical representation.

Further, the nature of the covariance functions used was explored. The squared covariance function, the neural

Table 3 List of the covariance functions applied in Gaussian process regression analysis

Covariance function	Description	Comments
Squared exponential covariance function	$k(\mathrm{x}_i, \mathrm{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathrm{x}_i - \mathrm{x}_j)^T M (\mathrm{x}_i - \mathrm{x}_j)\right)$	Where $M = l^{-2}I$, l is characteristic length-scale, and *I the identity matrix*. In these experiments, the *l is the same for each* feature (descriptor)
Neural network covariance function[a]	$k(\mathrm{x}_i, \mathrm{x}_j) = \alpha \sin^{-1}\left(\frac{\hat{x}_i^T \beta I \hat{x}_j}{\sqrt{(1+\hat{x}_i^T \beta I \hat{x}_i)(1+\hat{x}_j^T \beta I \hat{x}_j)}}\right)$	Where α and β are scalar hyperparameters to be optimized, $\hat{x}$ is the x vector extended by appending an element with the value 1. In these experiments, the neural network covariance function uses β as l shown above.
Rational quadratic covariance function	$k(\mathrm{x}_i, \mathrm{x}_j) = \left(1 + \frac{\lvert \mathrm{x}_i - \mathrm{x}_j \rvert^2}{2\alpha l^2}\right)^{-\alpha}$	Where α and l, the characteristic length-scale, are non-negative parameters of the covariance function.
Matern3 covariance function	polynomial order p; $p = 1$	Includes the *Matern* covariance functions with isotropic distance measures, where α and l, the characteristic length-scale, are non-negative parameters of the covariance function.
Matern5 covariance function	polynomial order p; $p = 2$	Includes the *Matern* covariance functions using l as above. The *Matern* covariance function can be defined as a product of an exponential and a polynomial of differing orders p (see middle column), where p is the highest exponent in the polynomial.

[a]Described by Neal (1996) as being equivalent to a feed-forward neural network with a single hidden layer (denoted as NN) in the limit of an infinite number of hidden units.[106]

network covariance function, the rational quadratic covariance function and two members of the *Matern* class of covariance function were explored.[93] An independent noise contribution was incorporated into the covariance function. These covariance functions are summarised in Table 3.

Performance measures

The measures used to characterise statistically model quality have been described in detail previously.[16,20,21,88] The parameters used are the *mean squared error* (MSE), *improvement over naïve model* (ION), *negative log estimated predictive density* (NLL) and the *Pearson correlation coefficient* (CORR). The MSE measures the averaged squared difference between model predictions and the corresponding targets. The ION measures the degree of improvement of the model over the naïve predictor (whose value is always the same; namely, the mean of experimental log K_p in the training set). The CORR measures the correlation between predictions and targets. For comparison, a model should aim to exhibit low values of both NLL and MSE and high values of ION and CORR for a given test dataset.

Analysis of the datasets

Firstly, the GP regression models for pig, rodent tissue and the artificial membranes were evaluated, and the most suitable covariance functions determined. GPs were applied with the different covariance functions described above and in Table 3 for each dataset. Further, the QSPR models were compared with GP models for the human skin dataset. For each different dataset analysis the 'leave-one-out' method was applied, and repeated for each member of the datasets in turn.

Secondly, given the number of animal experiments described in the literature, and the difficulties in obtaining human skin for experiments, the ability of a dataset consisting of permeability values from animal experiments (pig and rodent) to provide reasonable estimates of human skin permeability, particularly where current models of human skin have shown poor ability to model permeability (i.e. for compounds with a high log P value), was investigated by GP methods. The pig and rodent datasets were used separately as training sets and the trained GP model, using the neural network covariance function, was tested on the whole human skin dataset.

Thirdly, a quantitative comparison of human skin permeability predictions between GP-trained models on a rodent and human skin training dataset was undertaken. To make this comparison, the chemical compounds that were common to both datasets were used. There were 48 chemicals common to both of these datasets. A training set, which included the 48 common chemicals that had target values (i.e. experimentally measured values of K_p) in both the rodent and human skin datasets, was constructed. Two training models were then produced; one which predicted the skin permeability trained by using human skin data, and the other which was trained using rodent skin data. Previously 'unseen' human skin data (the remaining 92 compounds in the human skin dataset) was used as a test for both models. The GP model applied in the first experiment, described above, was again used (a GP model with five physicochemical descriptors and employing the *Matern3* covariance function).

Finally, the performance of a model trained on a combined dataset of rodent and human skin data was compared with that of a model trained solely on human skin, to assess the effect of such a combined dataset to model quality and predictivity. To avoid inconsistent training examples (i.e. examples with similar physicochemical descriptors but different target values), non-common chemical compounds were used as training examples. Thus, a human skin model was trained using human data (denoted as *trnH*) – the 92 non-common chemical compounds, and was tested using human data with the 48 common chemical compounds. To generate

a mixed model using human and rodent data, a mixed dataset was required. Further, a training set consisting of size 92 chemical compounds was used compare it with *trnH*, which had 92 compounds. Therefore, this was composed of chemical compounds that were not included in those common chemical compounds from the rodent dataset (55 compounds, denoted by *trnR*). Fifty samples were then randomly selected from *trnH*, and 42 from *trnR*, and added together and denoted as *trnHR*. The GP model was trained using *trnHR*. Finally, the GP model trained on *trnHR* was tested on the same human skin data as was used for the human skin model (i.e. 48 compounds). This procedure was repeated 10 times and the average results are shown in Table 5. Finally, the actual value of skin permeability (for rodent) was used as a prediction for human skin, as the test set contained only those compounds that were common to both rodent and human tissues.

Analysis was carried out via MatLab (R2008a) as described previously.[16] Briefly, the compounds were randomly allocated into the subsets automatically by Matlab via the *primeSeed* code. This acts as a recorder and documents the allocation of the compounds in the subsets. The experiment was repeated 10 times and generated 10 different training sets. Dealing with small datasets brings particular problems as test sets may be very small. Therefore, while there are many methods for dealing with this,[94] in this study we employed the simplest method available. Each training set contained a unique *primeSeed* code that recorded the compounds allocated in the corresponding training set. For the first three methods described above the leave-one-out method was employed, and for the final method randomly selected training sets, rather than the test set, were generated. Performance metrics were then computed in terms of all predictions.[20]

Results

Model validation

Table 4 lists the results of the model validation experiments, as described previously,[16] which describe the statistical quality of the models according to the performance measures described in the preceding section. For each experiment the best statistical outcome is highlighted in bold. It can be seen that GPs with different covariance functions outperform QSPRs on the human dataset and outperform the naïve model on all the datasets. It was observed that covariances *NN* and *Matern3* produced the best results, with regard to their statistical performance. If one example is randomly taken from the pig skin dataset (that compound with an experimentally determined log permeability coefficient of −6.29 cm/h, scopolamine, Table 1) and compared with GP predictions, the GP models with neural network and *Matern3* covariance functions yield predictions of, respectively, −5.79 and −4.47 cm/h. For the artificial membrane dataset, the GP model with *NN* covariance yielded the best performance, while the rational quadratic covariance function resulted in a worse performance than the naïve model. The other covariance methods produced results similar to the naïve model.

Table 4 Leave-one-out results for human, pig and rodent skin, and synthetic membrane datasets

Membrane	Model		MSE	ION	CORR	NLL
Human skin	QSPR	Moss & Cronin	20.09	−1157.60	0.14	–
		Potts & Guy	5.50	−244.53	0.10	–
	Naïve model		1.6	0	−1	–
	GP	NN	**1.13**	**29.14**	**0.53**	**1.48**
		SE	1.23	23.13	0.49	1.53
		RQ	1.13	29.06	0.53	9.42
		Matern3	1.20	25.08	0.51	9.43
		Matern5	1.21	24.26	0.50	9.43
Pig skin	Naïve model		2.50	0	−1	–
	GP	NN	0.59	76.52	0.86	1.65
		SE	0.64	74.45	0.84	20.65
		RQ	0.73	70.81	0.82	21.51
		Matern3	**0.51**	**79.74**	**0.88**	**1.08**
		Matern5	0.62	75.30	0.85	1.76
Rodent skin	Naïve model		1.30	0	−1	–
	GP	NN	0.88	32.63	0.56	1.41
		SE	0.86	34.39	0.58	1.40
		RQ	0.86	34.07	0.58	1.41
		Matern3	**0.83**	**36.25**	**0.60**	**1.38**
		Matern5	0.84	35.70	0.59	1.39
Synthetic membrane (polydimethylsiloxane, PDMS)	Naïve model		5.79	0	−1	–
	GP	NN	**3.57**	**38.26**	**0.60**	**2.03**
		SE	5.45	5.84	0.23	2.52
		RQ	6.33	−9.34	−0.70	2.96
		Matern3	5.55	4.15	0.08	2.72
		Matern5	5.19	10.37	0.22	2.65

CORR, Pearson correlation coefficient; GP, Gaussian Process; ION, improvement over naive model; MSE, mean squared error; NLL, negative log estimated predictive density or negative log loss; QSPR, quantitative structure-permeability relationship.

Absolute differences between predictions and experimental results

With regard to the accuracy of predictions for human skin, 87 out of 140 of the predictions produced by the GP model, with the neural network covariance function, were closer to the target (experimental) values than both the naïve model and the Potts and Guy 1992 QSPR model.[18] Further, 58 of the 103 predictions produced by the GP model were closer to the targets for the rodent data, compared with the naïve model. For this particular model, the comparison of predictions with experimental results appears to be dependent in some cases upon particular physicochemical descriptors, with the model performing better in certain ranges. For example, Figure 2 shows the differences between the predictions from the GP and naïve models, plotted against the log P values of the data, for porcine skin (Figure 2a) and the artificial membrane dataset (Figure 2b). It may be seen in these figures that the porcine skin model appears to give generally good predictions across a wide range of log P values, whereas the artificial membrane model appears to give uniformly poor performance for the same parameter. This indicates that the majority of the differences from GP models are smaller than those for the naïve model for the pig skin dataset. In the case of the artificial membrane dataset, nearly half the differences from GP modelling are smaller than those from the naïve model.

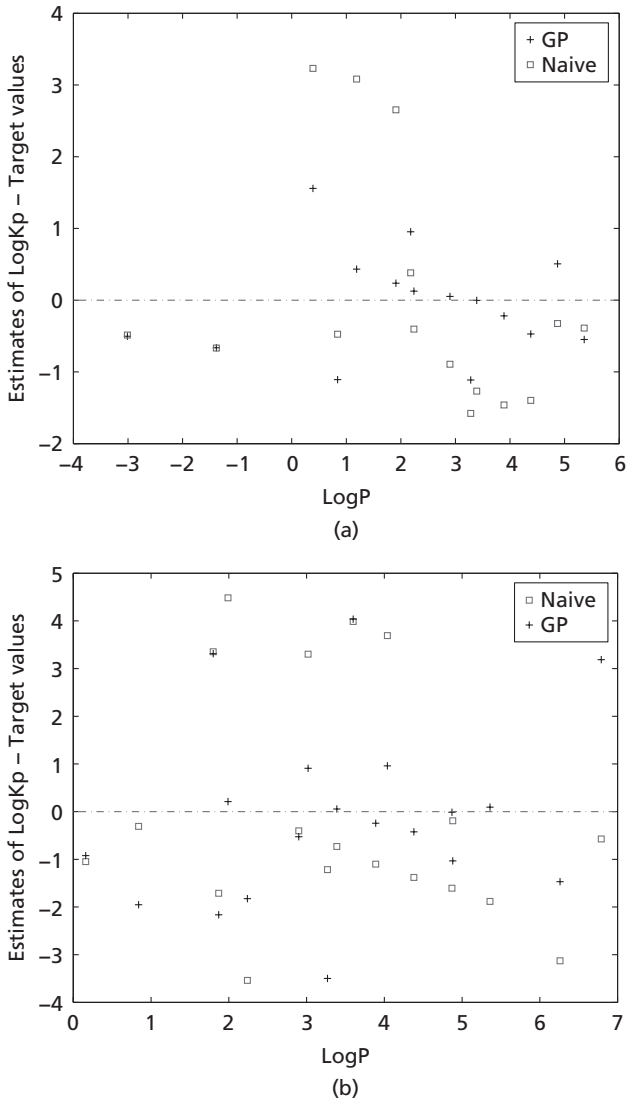

Figure 2 Differences between predictions and targets (experimental data) for (a) the pig skin and (b) polydimethylsiloxane (PDMS) datasets.

Comparison of animal and human data

The results of using the animal datasets to provide prediction of human skin permeation are shown in Table 5, with the best highlighted in bold. The results in Table 5 indicate that the GP models outperform the QSPR models, and that the model trained with the rodent dataset gives the best results in terms of accuracy of prediction. Further, the GP model trained using the pig dataset produced worse predictions than the naïve model, although this may be a result of the small size of the pig skin dataset. Figure 3 shows the experimental skin permeability coefficients obtained from human, pig and rodent datasets plotted against log P. It shows that examples from the pig dataset (Figure 3b) do not cover the whole range of physicochemical descriptors and target values of the human skin data (Figure 3a). Again, this may be due to the limited size of the pig skin dataset. In contrast, the relationship of rodent skin data to log P (Figure 3c) is reasonably similar to that with human skin data, which might suggest that the underlying mechanism of percutaneous absorption is similar for both species, and may indicate why the rodent skin dataset provides the best prediction of penetration into human skin.

Cross-comparison between datasets

Table 5 also shows the results from the comparison of human skin permeability predictions between GP-trained models on a rodent and human skin training dataset. The best results are again highlighted in bold. It can be seen that both the human and rodent training sets give better predictions on the human test set than using either the naïve model or QSPRs. Further, GP predictions from the rodent model are substantially better than the human skin naïve model.

Combining rodent and human data

The results of the addition of rodent skin data to a human skin training set are shown in Table 5. It can be seen that including rodent skin examples in the training set can produce predictions which, on average, are nearly as accurate as using a human training set of the same size. Finally, the results of using a GP model trained on rodent skin permeability values to predict human skin permeability are shown in Table 5. The results of this experiment are shown in the final row of Table 5, which looks at the statistical performance on the human test data set using models trained on the human, rodent and mixed training sets, respectively. The results in the final row of Table 5 indicate that they produce a prediction comparable to using human skin.

The results of this study show that, in general, GP methods produce better results, including better predictions of experimental targets, than QSPR models and naïve predictions when applied to the human and animal skin datasets employed in this work.

Table 5 Performance measures for the human skin dataset and those trained on the animal membrane datasets

Description of performance	Model		MSE	ION	CORR	NLL
Performance measured using the whole human skin dataset	QSPR	Moss & Cronin	20.09	−1157.60	0.14	–
		Potts & Guy	5.50	−244.53	0.10	–
	Pig skin	Naïve model	1.58	0	0	–
		GP-*Matern3*	1.70	−7.39	0.34	3.97
	Rodent skin	Naïve model	1.63	0	0	–
		GP-*Matern3*	**1.24**	**24.44**	**0.60**	**1.65**
Performances on the human test set using the model trained on rodent and human skin training sets, separately	QSPR	Moss & Cronin	19.35	−1203.3	0.16	–
		Potts & Guy	6.01	−304.59	0.12	–
	Human skin	Naïve model	1.48	0	0	–
		GP	**1.05**	**29.40**	**0.52**	**1.47**
	Rodent skin	Naïve model	1.37	0	0	–
		GP	1.13	17.22	0.46	1.47
Performance on the human test set using models trained on the human, rodent and mixed training sets, respectively.	Human skin	Naïve model	2.15	0	0	–
		GP	1.93	9.96	0.43	13.06
	Mixed	GP	1.91 ± 0.09	11.78 ± 3.05	0.51 ± 0.05	2.10 ± 0.31
	Rodent skin	GP	1.51	29.47	0.68	–

Note: QSPR ION results are from the human training set.
CORR, Pearson correlation coefficient; GP, Gaussian Process; ION, improvement over naive model MSE, mean squared error; NLL, negative log estimated predictive density or negative log loss; QSPR, quantitative structure-permeability relationship.

Discussion

The results of this study show that, in general, GP methods produce better results, including better predictions of experimental targets and statistical performance measures, than QSPR models and naïve predictions when applied to the human and animal skin datasets employed in this study. It also shows that better predictions of human skin permeability are obtained using models trained with rodent skin than with porcine skin, and that the former relates more closely to human skin absorption. They suggest that the GP models produce better results, both in a statistical sense, and in terms of the accuracy of prediction, than the quantitative structure-activity relationship (QSAR) models used to benchmark them. The results produced are, as in previous studies, consistent with an underlying non-linearity in the dataset.[16,20] This is particularly evident at the extremes of the model.

When Cronin and Schultz discussed the pitfalls of QSAR models, they listed a series of criteria for their use.[95] Key among their recommendations were the avoidance of extrapolation beyond the domain of the model, appreciation of the model's precision and its expected application in the context of the original biological measurement (in the case of percutaneous absorption, this relates to J_{max} or K_p), a single model that describes the whole process and not single steps or parts of the dataset employed, the avoidance of non-transparent QSARs and the correct understanding of the endpoint of the QSAR and its intended scope of use. Finally, they recommended the development of QSARs by multi-disciplinary groups of researchers whose expertise extends across all parts of the study and its methodology. Their work has been refined and has informed the OECD Principles for the Validation of (Q)SAR Models (available at http://www.oecd.org/document/4/0,3746,en_2649_34379_42926724_1_1_1_1,00.html). While this format provides a more robust description of QSARs for regulatory use, we have focused on the comments made by Cronin and Schultz.[95] While this manuscript, or the underlying Gaussian methods it uses, may ultimately have regulatory significance we believe that the key points this study makes, with regard to the construction and use of models, are more conceptual and, at this stage of the work, concern themselves more with the underlying principles and concepts of the Gaussian methods and model construction and development rather than validated models of regulatory significance.

Their latter point is particularly well emphasised in this, and previous studies,[16,20,21,88] and does impact on the issue of model transparency, an issue addressed and discussed in great detail previously, particularly in the use of feature selection methods by Lam *et al*.[16] To further reflect Cronin and Schultz's point on the construction of a multi-disciplinary team to examine this problem[95] this, and earlier, work[16,20,21,88] addressed their point in that a range of scientific disciplines contribute to this work. This particularly relates to the most fundamental work that underpins the use of GP models – the use of simple methods of data visualisation coupled with principal and canonical component analysis. This demonstrated the fundamental non-linearity of the dataset used for percutaneous absorption modelling, and highlights the issue with many studies that have adopted linear methods in their analyses.[15,20,21] It also reflects Flynn's comments which stated that very hydrophilic and very hydrophobic compounds were represented by different mathematical relationships, suggesting, among other things, non-linearity.[22] While these previous studies have been benchmarked against the Potts and Guy model,[18] due to its enormous contribution to this field and its widespread acceptance as the first quantitative model of percutaneous absorption, iterations of that model that include non-linear terms might provide more substantial and realistic benchmarks.

It is also important to build upon Cronin and Schultz's comment on the fundamental nature of a QSAR or other 'mathematical model' of a biological process, which they define as a triangulation of biological endpoint data, physico-

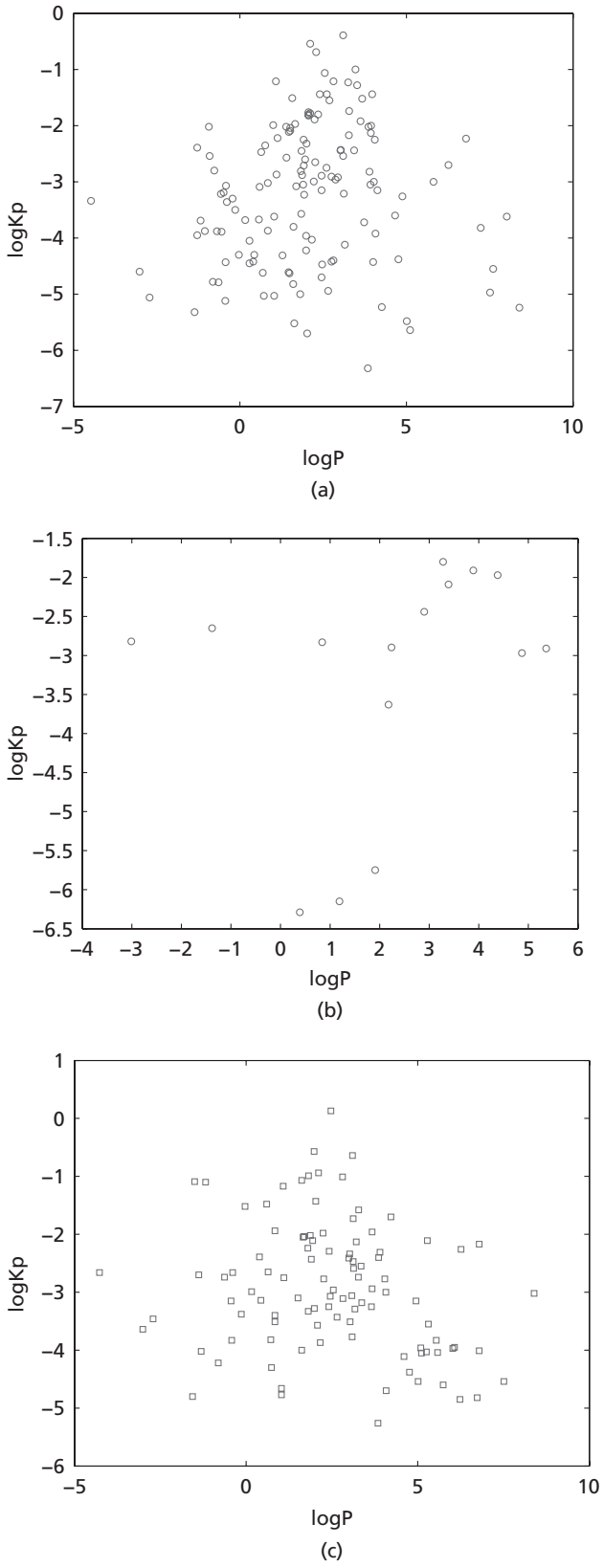

Figure 3 Permeability coefficients as a function of log P for the (a) human (b) pig and (c) rodent datasets.

chemical and structural information of the chemicals of which the dataset is comprised and a suitable mathematical and/or statistical approach to model development.[95] This might suggest that the nature of the experimental data (in this case, the K_p and J_{max} data derived from laboratory-based experiments) should not be removed from the interpretation or use of the final model produced.

One point not directly raised by Cronin and Schultz is the quality of the data used to construct the model. This was reviewed by Moss *et al*.[15] who commented upon the wide range of methods used by researchers and how this impacts on the quality and variance observed in their datasets. Inevitably, the nature of the biological membrane used clearly has an important role in the model that is constructed from this data. This is well understood and discussed elsewhere.[15] However, one issue that is not as widely discussed is the number of data points required to produce a valid model.

The merits, or otherwise, of extrapolating a model substantially beyond its boundaries (defined by the range of data input to the model) is both common sense and a clear limitation in the applicability of the model. It might also be suggested that, if the model is developed from a dataset that is poorly or unevenly distributed – possibly exhibiting a high degree of redundancy in its data – that it might also be difficult to develop a model that is uniformly accurate even within these boundaries. Sun *et al*.[21] demonstrated substantial increases in covariance as the GP models extended outside those areas of the dataset that were highly populated. While this also implies criticism of comparisons made between GP models and, for example, the Potts and Guy equation,[18] it is important to demonstrate the fallibility of models in such circumstances to avoid inappropriate use. This was highlighted by Moss *et al*.[20,96] who showed how QSAR models of percutaneous absorption fail to accurately predict the permeability coefficient, K_p, across a range of log P values. The authors of that study also clearly demonstrated the general failure of prediction across the whole range under examination. In the current study the ranges of data (i.e. for the parameters investigated) used for human, porcine, rat and artificial membranes were broadly comparable. More generally, the failure of models to predict a wide range of permeability (in terms of physicochemical properties) suggests that the models developed are quite poor in terms of their applicability. For example, at very low or very high log P values, metals and excipients used in pharmaceutical and cosmetic formulations are poorly modelled and their absorption is quite poorly understood. Reinforcing this limitation is the preponderance of linear models used in this field. Models such as the Potts and Guy equation[18] poorly fit the extremes of physicochemical properties described above, but would most likely be unable to describe them anyway due to the underpinning methodology used to generate the models. The GP approach also allows the use of a single model to cover the whole range of the data being examined, and does not rely on modifications to older models to try and expand the range of applicability. It should also be noted that the models developed previously[20,88] are based on datasets that are approximately 50% larger than those of, for example, Potts and Guy.[18] Hence, while modelling outside the ranges of a dataset is not desirable, the GP method appears to be able to accommodate

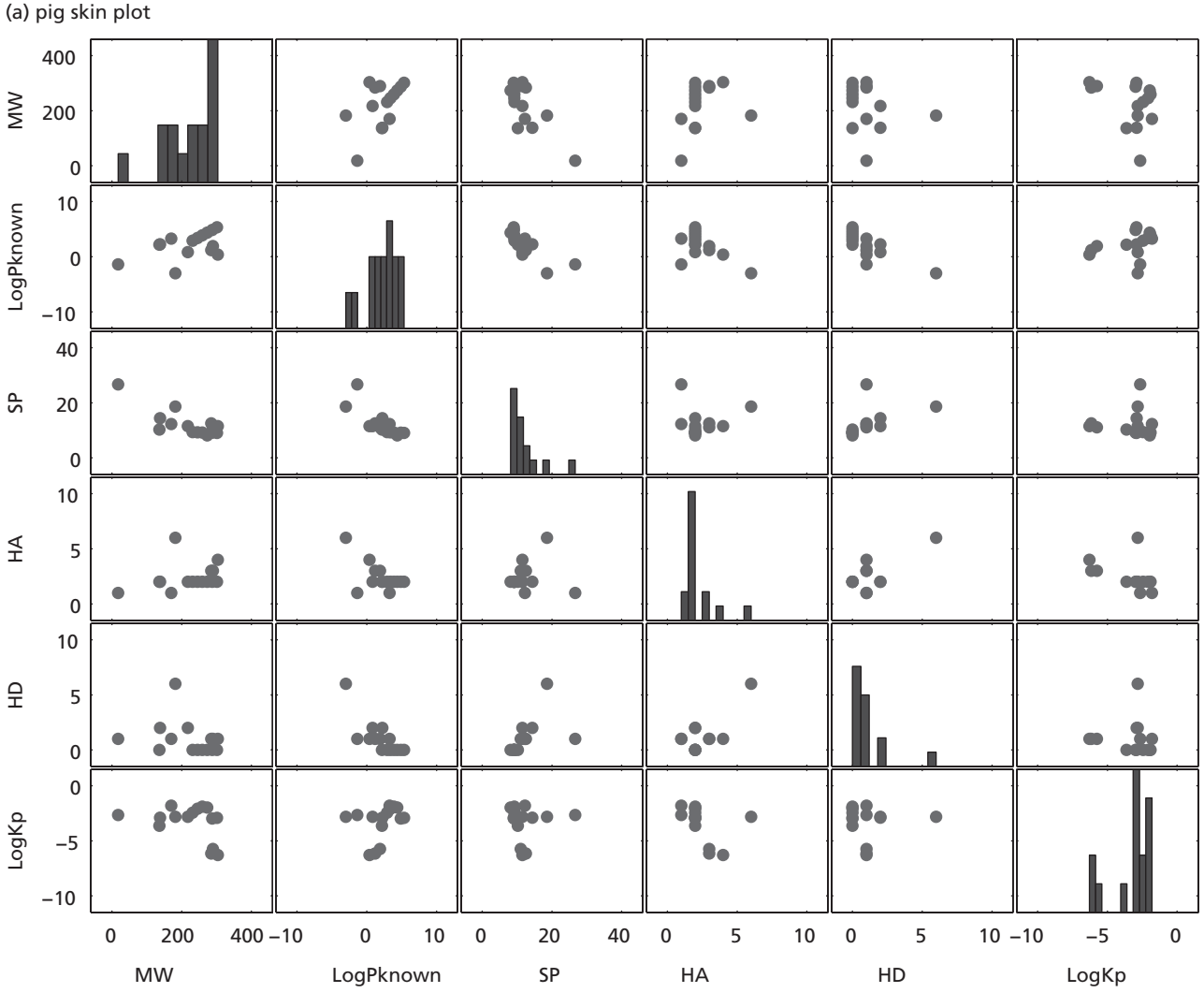

Figure 4 Scatter plots for the pig and rodent datasets. (a) Pig skin plot. (b) Rodent skin plot.

such data points better than linear models. Such discussions inevitably lead on to the subject of outliers, which has been described elsewhere.[15,95]

The human, pig, rodent and artificial membrane datasets consist of 140, 15, 103 and 19 chemical compounds, respectively, making the human skin dataset in particular one of the largest datasets used for the modelling of human skin absorption. In addition, the dataset for the rodent skin is also relatively large, with 103 members. Of equal importance are the small datasets for artificial membranes and porcine skin, with 19 and 15 compounds in each dataset, respectively. Clearly, the nature of the porcine skin dataset – most notably its size and the inherent range of experimental protocols used in those original studies – significantly impacts on not just the quality of the porcine skin GP model, but it also puts the findings of the other models into context. For example, the clear 'false positive' of this study is that the rodent skin dataset produces a more accurate model than porcine skin. While this would suggest that rodent skin is more representative of human skin than porcine skin, we are aware of the vast body of literature, from Wester and Noonan[1] onwards, that presents experimental evidence that this is not the case. It may also be the case that re-enforcement of model 'quality' – good or bad – occurs with an increase in randomly selected data. While specific data selection is somewhat subjective it is reasonable to suggest that a spread of data evenly throughout the dataset might remove issues of skew and possible bias in the development of the QSAR or GP models. Recently, Sun *et al*.[21] visualised the distribution of a human skin dataset ($n = 149$) and showed that the data was unevenly distributed, based on a number of physicochemical parameters, across the whole dataset. In the current study, the visualisation and distribution of the datasets are shown in Figure 4. This clearly shows some bias in certain parts of the dataset (as shown on the histograms, running top left to bottom right, in Figure 4a and 4b). This may suggest that, while a particular dataset is quite large in size, an uneven distribution of data points may affect the model produced; hence, the size of the dataset may only be relevant when it is evenly distributed across as wide a range of physicochemical properties as possible.

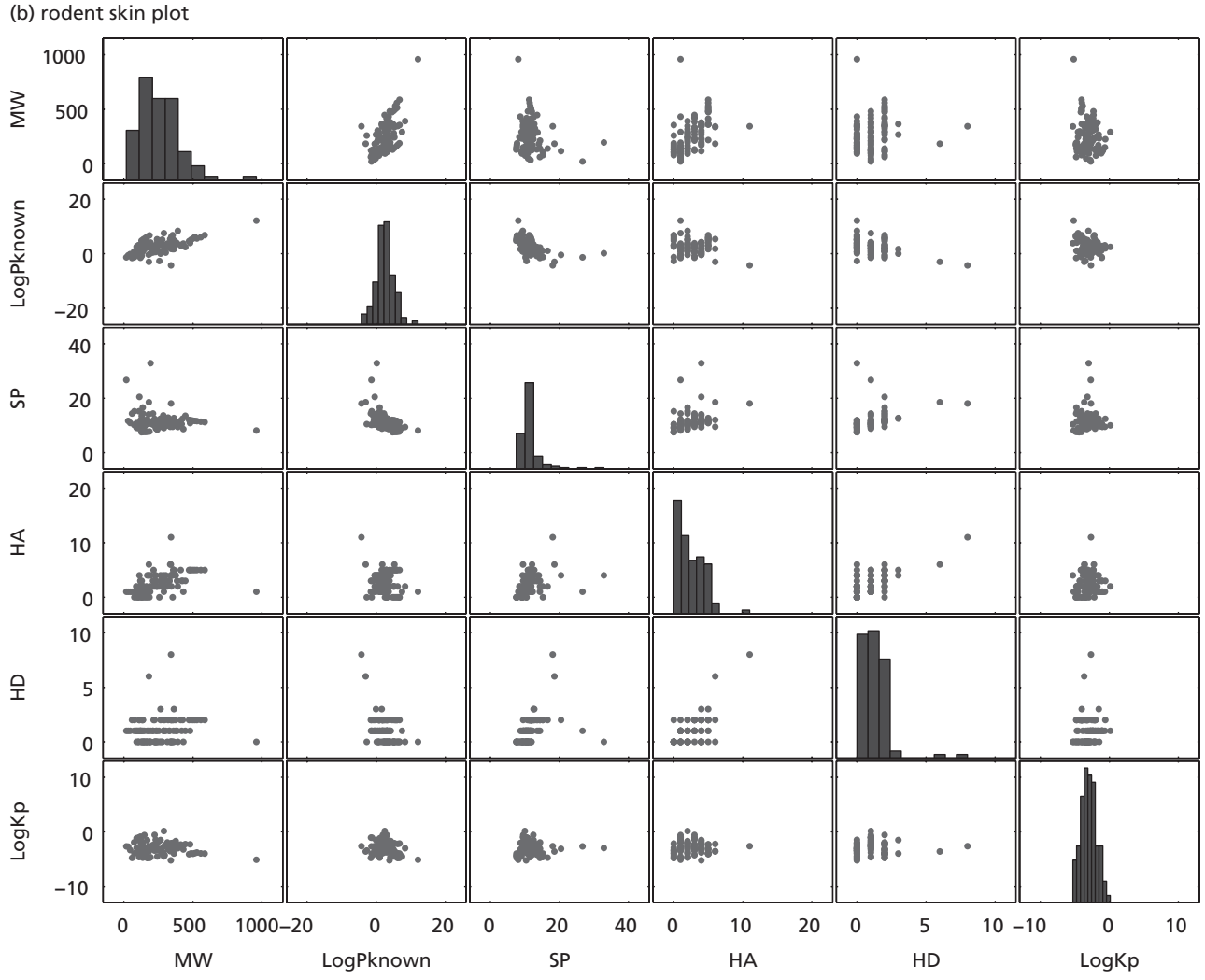

Figure 4 *(Continued)*

Such an understanding of both the size and distribution of the data used to develop models might impact on the quality of analysis. For example, Lien and Gao[97] analysed a subset of the Flynn dataset[22] which comprised twenty-two compounds ($r^2 = 0.96$). Other studies[84] have, for various reasons, examined only non-electrolytes from the Flynn dataset ($n = 37$, $r^2 = 0.94$), selected subsets of a larger database or examined a small number of compounds and derived QSAR-type models. These range from the work of Barratt,[98] which examined 60 'small molecules and steroids' ($n = 60$; $r^2 = 0.90$) excluding the hydrocortisone derivatives, from the Flynn dataset, Abrahams *et al*.,[99,100] who examined, respectively, 46 and 53 compounds, to those who analysed substantially smaller datasets with 20,[101] 16 (in two different studies by both Lee *et al*.[102] and Morimoto *et al*.[103]) or four compounds.[104] In all these cases, mechanistic inferences were drawn into the percutaneous absorption of these compounds and, by inference, those that are chemically similar. However, given the issues presented in this study with the accuracy of small datasets, and their examination by GP methods that have recently been shown to exhibit superior statistical and predictive accuracy compared with a number of QSAR-type models, the number of compounds present in these datasets (despite claims of 'diverse' datasets, by Morimoto *et al*.[103]) would suggest that the value of those models may be limited by the amount of available data. The work of Potts and Guy[18,84] raises an important issue in this matter; their models are substantially different, as the first model[18] focuses on the whole Flynn dataset,[84] whereas the second[84] used only the 37 non-electrolytes from that dataset. Further, the high correlation coefficients for some of the models derived from small datasets might suggest over-fitting of the data, or that the selection of a particular subset presents an unrealistic representation of the statistical nature of those models. The inclusion of a deliberately small dataset in this study clearly shows the impact the volume of data can have on the quality of the model, and presents quite clearly the possibility of developing a misleading model, which reflects the comments on model design and quality by Cronin and Schultz.[95]

Cronin and Schultz commented that biological processes are seldom linear in nature and that 'global modelling is unlikely to be successful without the consideration of

non-linearity'.[95] However, over-modelling may result in simply modelling the error present in the data. This may be an issue when considering the GP model produced for rodent skin in this study, particularly with regard to the distribution of the physicochemical properties of the data. For example, while the rodent GP model is statistically better than the porcine model, the preceding section has described that the majority of predictions made with the human skin data are more precise than those made with the rodent data (i.e. Figure 2 and Table 4). Therefore, while the rodent dataset provides a better model – in these experiments – than the porcine skin model, it is still not as accurate as the human skin model (Table 4). Clearly, a larger porcine skin dataset might improve the accuracy of that model, relative to the rodent skin model.

The results of this study indicate that permeation across two animal membranes (rodent and pig skin) is, in a statistical sense, similar. It also suggests that the artificial membranes are poor replacements for human or animal skin (Table 4). However, the overriding issue raised in this study is the nature of a dataset and how it can influence the results, and subsequent interpretation, of any model produced. The size of the datasets, in both absolute and comparative senses, appears to influence model quality, producing counterintuitive results that simply do not agree with a large literature of laboratory-generated experimental data (i.e., that rat skin can be used to predict human skin absorption, or that rat skin is a 'better', or more accurate, model than porcine skin for human skin in permeability experiments).

Further, while it may seem appropriate to collate datasets together and produce larger datasets, such collations may not necessarily result in better models, given the underlying distribution of the data and the physicochemical parameters represented. Hence, an efficiency in dataset design, including an even and representative spread of data, avoiding data redundancy, may be more important to dataset quality than simply adding all the new percutaneous absorption data as it appears in the literature. Such an argument does not even begin to address issues of experimental design, as highlighted by Moss *et al*.;[15] for example, the use of both experimental and calculated (often by several different means) log P values in datasets.[96] It would also suggest that, if a model is to be generated using literature data from experiments based on rodent or porcine skin, such a model would need to be validated thoroughly as, in the presence of substantial numbers of experiments on human skin, such a model would be redundant for any predictive purpose, but might yield important mechanistic differences. However, the quality of the dataset would need to improve substantially from that presented in this study (particularly for porcine skin) if it was to be of any realistic use.

Conclusions

The over-riding issue addressed in this study is that the development of models for percutaneous absorption, for any species, is very much dependent on the nature and quality of the data used to construct the model, and that Cronin and Schultz's comments on model development should be considered when constructing mathematical representations of percutaneous absorption to avoid the generation of false positive or false negative results.[95] This clearly reflects comments made elsewhere, including by Cronin and Schultz and subsequent OECD criteria for the validation of QSARs. It is also reflected in, for example, the work of Buist *et al*.,[105] who developed models for finite dose experiments. While this clearly has significance for understanding skin absorption and is an important study, it also highlights the issues of data availability and dataset construction discussed herein, as not only does it focus on finite dose studies but also it is only relevant for non-volatile compounds.

Declarations

Conflict of interest

The Author(s) declare(s) that they have no conflicts of interest to disclose.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

1. Wester RC, Noonan PK. Relevance of animal models for percutaneous absorption. *Int J Pharm* 1980; 7: 99–110.
2. Williams AC. *Transdermal and Topical Drug Delivery*. London: Pharmaceutical Press, 2003: 56.
3. Washington C. Drug release from microdisperse systems – a critical review. *Int J Pharm* 1980; 58: 1–12.
4. Flynn GL *et al*. Mass transport phenomena and models: theoretical concepts. *J Pharm Sci* 1974; 63: 479–510.
5. Garrett ER, Chemburkar PB. Evaluation, control and prediction of drug diffusion through polymeric membranes II. *J Pharm Sci* 1968; 57: 949–959.
6. Nacht S, Yeung D. Artificial membranes and skin permeability. In: Bronaugh RL, Maibach HI, eds. *Percutaneous Absorption; Mechanisms – Methodology – Drug Delivery*. New York: Marcel Dekker, 1985: 373.
7. Kurosaki Y *et al*. Use of lipid disperse systems in transdermal drug delivery – comparative study of flufenamic acid permeation among rat abdominal skin, silicon rubber membrane and stratum corneum sheet isolated from hamster cheek pouch. *Int J Pharm* 1991; 67: 1–9.
8. Megrab NA *et al*. Estradiol permeation through human skin and Silastic membrane – effects of propylene glycol and supersaturation. *J Controlled Release* 1995; 36: 277–294.
9. Megrab NA *et al*. Estradiol permeation through human skin, Silastic and snake skin membranes – the effects of ethanol-water co-solvent systems. *Int J Pharm* 1995; 116: 101–112.
10. Stott PW *et al*. Characterisation of complex coacervates of some tricyclic antidepressants and evaluation of their potential for enhancing transdermal flux. *J Controlled Release* 1996; 41: 215–227.
11. Woolfson AD *et al*. Development and characterisation of a moisture-activated bioadhesive drug delivery system for percutaneous local anaesthesia. *Int J Pharm* 1998; 169: 83–94.
12. Minghetti P *et al*. Comparison of different membranes with cultures of keratinocytes from man for percutaneous absorption of nitroglycerine. *J Pharm Pharmacol* 1999; 51: 673–678.
13. Moss GP *et al*. Design, synthesis and characterisation of captopril prodrugs for enhanced percutaneous absorption. *J Pharm Pharmacol* 2006; 58: 167–177.

14. Cronin MTD *et al*. An investigation of the mechanism of flux across polydimethylsiloxane membranes by use of quantitative structure-permeability relationships. *J Pharm Pharmacol* 1998; 50: 143–152.
15. Moss GP *et al*. Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicol In Vitro* 2002; 16: 299–317.
16. Lam LT *et al*. The application of feature selection to the development of Gaussian process models for percutaneous absorption. *J Pharm Pharmacol* 2010; 62: 738–749.
17. Bronaugh RL *et al*. Methods for *in vitro* percutaneous absorption studies I: comparison with *in vivo* results. *Appl Pharmacol* 1982; 62: 481–488.
18. Potts RO, Guy RH. Predicting skin permeability. *Pharm Res* 1992; 9: 663–669.
19. Geinoz S *et al*. Quantitative structure–permeation relationships (QSPeRs) to predict skin permeation: a critical evaluation. *Pharm Res* 2004; 21: 83–92.
20. Moss GP *et al*. The application of Gaussian processes to the prediction of percutaneous absorption. *J Pharm Pharmacol* 2009; 61: 1147–1153.
21. Sun Y *et al*. The application of stochastic machine learning methods in the prediction of skin penetration. *Appl Soft Comput* 2011; 11: 2367–2375.
22. Flynn GL. Physicochemical determinants of skin absorption. In: Gerrity TR, Henry CJ, eds. *Principles of Route-to-Route Extrapolation for Risk Assessment*. New York: Elsevier, 1990: 93–127.
23. Cnubben NHP *et al*. Comparative *in vitro* and *in vivo* percutaneous penetration of the fungicide *ortho*-phenylphenol. *Regul Toxicol Pharmacol* 2002; 35: 198–208.
24. Bosman IJ *et al*. Standardisation procedure for the *in vitro* skin permeation of anticholinergics. *Int J Pharm* 1998; 169: 65–73.
25. Zhang C *et al*. Effects of cinnamene enhancers on transdermal delivery of ligustrazine hydrochloride. *Eur J Pharm Biopharm* 2007; 67: 413–419.
26. Clowes HM *et al*. Skin absorption: flow-through or static diffusion cells. *Toxicol In Vitro* 1994; 8: 827–830.
27. Schmook FP *et al*. Comparison of human skin or epidermis models with human and animal skin in *in vitro* percutaneous absorption. *Int J Pharm* 2001; 215: 51–56.
28. Tojo K *et al*. Drug permeation across the skin: effect of penetrant hydrophilicity. *J Pharm Sci* 2006; 76: 123–126.
29. McDougal JN *et al*. Assessment of skin absorption and penetration of JP-8 jet fuel and its components. *Toxicol Sci* 2000; 55: 247–255.
30. Moody RP, Nadeau B. In vitro dermal absorption of two commercial formulations of 2,4-dichlorophenoxyacetic acid dimethylamine (2,4-D amine) in rat, guinea pig and human skin. *Toxicol In Vitro* 1997; 11: 251–262.
31. Lockley DJ *et al*. Percutaneous penetration and metabolism of 2-ethoxyethanol. *Toxicol Appl Pharmacol* 2002; 180: 74–82.
32. Dies-Sales O *et al*. The prediction of percutaneous absorption: I. Influence of the dermis on in vitro permeation models. *Int J Pharm* 1993; 100: 1–7.
33. Brand RM, Mueller C. Transdermal penetration of atrazine, alachlor and trifluralin: effect of formulation. *Toxicol Sci* 2002; 68: 18–23.
34. Calpena AC *et al*. A comparative in vitro study of transdermal absorption of antiemetics. *J Pharm Sci* 2006; 83: 29–33.
35. Fang L *et al*. The use of complexation with alkanolamines to facilitate skin permeation of mefenamic acid. *Int J Pharm* 2003; 262: 13–22.
36. Hotchkiss SA *et al*. Percutaneous absorption of benzyl acetate through rat skin in vitro. 1. Validation of an in vitro model against in vivo data. *Food Chem Toxicol* 1990; 28: 443–447.
37. Boogaard PJ *et al*. Dermal penetration and metabolism of five glycidyl ethers in human, rat and mouse skin. *Xenobiotica* 2000; 30: 469–483.
38. Simonsen L *et al*. In vivo skin penetration of salicylic compounds in hairless rats. *Eur J Pharm Sci* 2002; 17: 95–104.
39. Greaves LC *et al*. Factors affecting the percutaneous absorption of caffeine in vitro. *Toxicology* 2002; 178: 65–66.
40. Beckley-Kartey SA *et al*. Comparative in vitro skin absorption and metabolism of coumarin (1,2-benzopyrone) in human, rat and mouse. *Toxicol Appl Pharmacol* 1997; 145: 34–42.
41. Moody RP *et al*. *In vitro* dermal absorption of pesticides: VI. *In vivo* and *in vitro* comparison of the organochlorine insecticide DDT in rat, guinea pig, pig, human and tissue-cultured skin. *Toxicol In Vitro* 1994; 8: 1225–1232.
42. Hughes MF *et al*. *In vitro* dermal absorption of flame retardant chemicals. *Food Chem Toxicol* 2001; 39: 1263–1270.
43. Pelling D *et al*. Absorption of hydrophilic and lipophilic compounds through epidermal and subepidermal strata of rat skin *in vitro*. *Toxicol In Vitro* 1997; 12: 47–55.
44. Scott RC *et al*. *In vitro* absorption of some *o*-phthalate diesters through human and rat skin. *Environ Health Perspect* 1987; 74: 223–227.
45. Sherertz EF, Sloan KB. Percutaneous penetration of squaric acid and its esters in hairless mouse and human skin *in vitro*. *Arch Dermatol Res* 1988; 280: 57–60.
46. Hall LL *et al*. Age-related percutaneous penetration of 2-sec-butyl-4,6-dinotrophenol (Dinoseb) in rats. *Toxicol Sci* 1992; 19: 258–267.
47. Scott RC *et al*. The influence of skin structure on permeability – an intersite and interspecies comparison with hydrophilic penetrants. *J Invest Dermatol* 1991; 96: 921–925.
48. Jolicoeur LM *et al*. Etorphine is an opiate analgesic physicochemically suited to transdermal delivery. *Pharm Res* 1992; 9: 963–965.
49. Diez I *et al*. A comparative *in vitro* study of transdermal absorption of a series of calcium-channel antagonists. *J Pharm Sci* 1991; 80: 931–934.
50. Moody RP, Ritter L. An automated *in vitro* dermal absorption procedure. 2. Comparative *in vivo* and *in vitro* dermal absorption of the herbicide fenoxaprop-ethyl (HOE-33171) in rats. *Toxicol In Vitro* 1992; 6: 53–59.
51. Barber ED *et al*. The percutaneous absorption of hydroquinone (HQ) through rat and human skin *in vitro*. *Toxicol Lett* 1995; 80: 167–172.
52. Kushla GP, Zatz JL. Influence of pH on lidocaine penetration through human and hairless mouse skin *in vitro*. *Int J Pharm* 1991; 71: 167–173.
53. Hoelgaard A, Mollgaard B. Permeation of linoleic acid through skin *in vitro*. *J Pharm Pharmacol* 1982; 34: 610–611.
54. Dick IP, Scott RC. The influence of different strains and age on *in vitro* rat skin permeability to water and mannitol. *Pharm Res* 1992; 9: 884–887.
55. Behl CR *et al*. Hydration and percutaneous absorption 1. Influence of hydration on alkanol permeation through hairless mouse skin. *J Invest Dermatol* 1980; 75: 346–352.
56. Dal Pozzo A *et al*. Percutaneous absorption of nicotinic acid derivatives *in vitro*. *J Pharm Sci* 1991; 80: 54–57.
57. Monti D *et al*. Comparison of ultrasound and of chemical enhancers on transdermal penetration of caffeine and morphine through hairless mouse skin *in vitro*. *Int J Pharm* 2001; 229: 131–137.

58. Sato K *et al*. Species differences in percutaneous absorption of nicorandil. *J Pharm Sci* 1991; 80: 104–107.
59. Moody RP, Nadeau B. An automated *in vitro* dermal absorption procedure. 3. *In vivo* and *in vitro* comparison with the insect repellent N,N-diethyl-*m*-toluamide in mouse, rat, guinea pig, pig, human and tissue-cultured skin. *Toxicol In Vitro* 1993; 7: 167–176.
60. Hashiguchi T *et al*. *In vitro* percutaneous absorption of prednisolone derivatives based on solubility parameter. *Int J Pharm* 1997; 158: 11–18.
61. Kunta JR *et al*. Effect of menthol and related terpenes on the percutaneous absorption of propranolol across excised hairless mouse skin. *J Pharm Sci* 1997; 86: 1369–1373.
62. Harada K *et al*. In vitro permeability to salicylic acid of human, rodent and shed snake skin. *J Pharm Pharmacol* 1993; 45: 414–418.
63. Tsai JC *et al*. Effect of barrier disruption by acetone treatment on the permeability of compounds with various lipophilicities: implications for the permeability. *J Pharm Sci* 2001; 90: 1242–1254.
64. Twist JN, Zatz JL. The effect of solvents on solute penetration through fuzzy rat skin *in vitro*. *J Soc Cosmet Chem* 1989; 40: 231–242.
65. Ackermann C, Flynn GL. Ether water partitioning and permeability through nude mouse skin *in vitro*. 1. Urea, thiourea, glycerol and glucose. *Int J Pharm* 1987; 36: 61–66.
66. Moss T *et al*. Percutaneous penetration and dermal metabolism of triclosan (2,4,4'-trichloro-2'-hydroxydiphenyl ether). *Food Chem Toxicol* 2000; 38: 361–370.
67. Bronaugh RL *et al*. Methods for *in vitro* percutaneous absorption studies. 1. Comparison with *in vivo* results. *Toxicol Appl Pharmacol* 1982; 62: 474–480.
68. Van de Sandt JJM *et al*. Comparative in vitro in vivo percutaneous absorption of the pesticide propoxur. *Toxicol Sci* 2000; 58: 15–22.
69. Kim MK *et al*. Skin permeation of testosterone and its ester derivatives in rats. *J Pharm Pharmacol* 2000; 52: 369–375.
70. Roper CS *et al*. Percutaneous penetration of 2-phenoxyethanol through rat and human skin. *Food Chem Toxicol* 1997; 35: 1009–1016.
71. Rigg PC, Barry BW. Shed snake skin and hairless mouse skin as model membranes for human skin during permeation studies. *J Invest Dermatol* 1990; 94: 235–240.
72. Jain GK *et al*. In vitro transdermal delivery of atenolol using mouse and guinea pig. *Indian J Exp Biol* 1993; 31: 691–693.
73. Bronaugh RL *et al*. Differences in permeability of rat skin related to sex and body site. *J Soc Cosmet Chem* 1983; 34: 127–135.
74. Vaddi HK *et al*. Effect of some enhancers on the permeation of haloperidol through rat skin *in vitro*. *Int J Pharm* 2001; 212: 247–255.
75. Behl CR *et al*. Percutaneous absorption of corticosteroids – age, site and skin-sectioning influences on rates of permeation of hairless mouse skin by hydrocortisone. *J Pharm Sci* 1984; 73: 1287–1290.
76. Nokhodchi A *et al*. The enhancement effect of surfactants on the penetration of lorazepam through rat skin. *Int J Pharm* 2003; 250: 359–369.
77. Scott RC *et al*, ed. *Prediction of Percutaneous Penetration: Methods, Measurements, Modelling*. London: IBC Technical Services Ltd, ISBN 1 85271 117 5. 1990.
78. Gay R *et al*. The living skin equivalent as a model in vitro for ranking the toxic potential of dermal irritants. *Toxicol In Vitro* 1992; 6: 303–315.
79. Moody RP *et al*. *In vitro* dermal absorption of pesticides: V. In vivo and in vitro comparison of the herbicide 2,4-dichlorophenoxyacetic acid in rat, guinea pig, pig, human and tissue-cultured skin. *Toxicol In Vitro* 1994; 8: 1219–1224.
80. Wagner H *et al*. Interrelation of permeation and penetration parameters obtained from *in vitro* experiments with human skin and skin equivalents. *J Control Release* 2001; 75: 283–295.
81. Ernesti AM *et al*. Absorption and metabolism of topically applied testosterone in an organotypic skin culture. *Skin Pharmacol Physiol* 1992; 5: 146–153.
82. Roberts M. Percutaneous absorption of phenolic compounds. Sydney, Australia: University of Sydney, 1976 (dissertation).
83. Kamlet MJ *et al*. Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^*, α, and β, and some methods for simplifying the generalized solvatochromic equation. *J Org Chem* 1983; 48: 2877–2887.
84. Potts RO, Guy RH. A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. *Pharm Res* 1995; 12: 1628–1633.
85. Roberts M *et al*. Epidermal permeability-penetrant structure relationships: 1. An analysis of methods of predicting penetration of monofunctional solutes from aqueous solutions. *Int J Pharm* 1995; 126: 219–233.
86. Roberts M *et al*. Epidermal permeability: penetrant structure relationships. 2. The effect of H-bonding groups in penetrants on their diffusion through the *stratum corneum*. *Int J Pharm* 1996; 132: 23–32.
87. Pugh WJ *et al*. Epidermal permeability – penetrant structure relationships: 3. The effect of hydrogen bonding interactions and molecular size on diffusion across the *stratum corneum*. *Int J Pharm* 1996; 138: 149–165.
88. Sun Y *et al*., Predictions of Skin Penetration Using Machine Learning Methods. In: Giannotti F *et al*., ed. *Proceedings of 8th IEEE International Conference on Data Mining, Pisa, Italy*. Washington, DC: IEEE Computer Society, 2008: 1049–1054. ISBN 978-0-7695-3502-9.
89. Fedors RF. A method for estimating both the solubility parameters and molar volumes of liquids. *Polym Eng Sci* 1974; 14: 147–154.
90. Moss GP, Cronin MTD. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. *Int J Pharm* 2002; 238: 105–109.
91. Johnson ME *et al*. Permeation of steroids through human skin. *J Pharm Sci* 1995; 84: 1144–1146.
92. Lin RY *et al*. A method to predict the transdermal permeability of amino acids and dipeptides through porcine skin. *J Controlled Release* 1996; 38: 229–234.
93. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press, 2006.
94. Golbraikh A, Tropsha A. A QSAR modelling using chirality descriptors derived from molecular topology. *J Chem Inf Comput Sci* 2002; 16: 357–369.
95. Cronin MTD, Schultz TW. Pitfalls in QSAR. *J Mol Struct* 2003; 622: 39–51.
96. Moss GP *et al*. A comparison of predicted and experimentally measured descriptors in QSPRs. *J Pharm Pharmacol* 2006; 58S: 104.
97. 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: 583–587.
98. Barratt MD. Quantitative structure–activity relationships for skin permeability. *Toxicol In Vitro* 1995; 9: 27–37.
99. Abraham MH *et al*. The factors that influence skin penetration of solutes. *J Pharm Pharmacol* 1995; 47: 8–16.

100. Abraham MH *et al*. Hydrogen bonding part 46. A review of the correlation and prediction of transport properties by an LFER method: physicochemical properties, brain penetration and skin permeability. *Pestic Sci* 1999; 55: 78–88.
101. Hostynek JJ, Magee PS. Modelling in vivo human skin absorption. *Quant Struct Act Relat* 1997; 16: 473–479.
102. Lee CK *et al*. Skin permeability of various drugs with different lipophilicity. *J Pharm Sci* 1994; 4: 562–565.
103. Morimoto Y *et al*. Prediction of skin permeability of drugs: comparison of human and hairless rat skin. *J Pharm Pharmacol* 1992; 44: 634–639.
104. Kitagawa S *et al*. Skin permeation of parabens in excised guinea pig dorsal skin, its modification by penetration enhancers and their relationship with n-octanol/water partition coefficients. *Chem Pharm Bull* 1997; 45: 1354–1357.
105. Buist HE *et al*. New *in vitro* dermal absorption database and the prediction of dermal absorption under finite conditions for risk assessment purposes. *Regul Toxicol Pharmacol* 2010; 57: 200–209.
106. Neal RM. Bayesian Learning for Neural Networks. Lecture Notes in Statistics No. 118. New York: Springer-Verlag, 1996.