JPP JOURNAL OF Pharmacy and Pharmacology
ROYAL PHARMACEUTICAL SOCIETY
JPP Journal of Pharmacy And Pharmacology
Research Paper

An evaluation of the potential of linear and nonlinear skin permeation models for the prediction of experimentally measured percutaneous drug absorption

Marc B. Brown[a,b], Chi-Hian Lau[a], Sian T. Lim[a], Yi Sun[b], Neail Davey[b], Gary P. Moss[c], Seon-Hie Yoo[d] and Christian De Muynck[d]

[a]MedPharm Ltd, Unit 3/Chancellor Court, Surrey Research Park, Guildford, [b]School of Pharmacy, University of Hertfordshire, College Lane, Hatfield, [c]School of Pharmacy, Keele University, Keele, UK, and [d]Nycomed, Byk-Gulden-Str. 2, Konstanz, Germany

Keywords
in-silico; in-vitro; mathematical models; prediction; skin permeation

Correspondence
Marc B. Brown, MedPharm Ltd, Unit 3/Chancellor Court, 50 Occam Road, Surrey Research Park, Guildford, GU2 7AB, UK.
E-mail: marc.brown@medpharm.co.uk

Received July 26, 2011
Accepted November 21, 2011

doi: 10.1111/j.2042-7158.2011.01436.x

Abstract

Objectives The developments in combinatorial chemistry have led to a rapid increase in drug design and discovery and, ultimately, the production of many potential molecules that require evaluation. Hence, there has been much interest in the use of mathematical models to predict dermal absorption. Therefore, the aim of this study was to test the performance of both linear and nonlinear models to predict the skin permeation of a series of 11 compounds.

Methods The modelling in this study was carried out by the application of both quantitative structure permeability relationships and Gaussian process-based machine learning methods to predict the flux and permeability coefficient of the 11 compounds. The actual permeation of these compounds across human skin was measured using Franz cells and a standard protocol with high performance liquid chromatography analysis. Statistical comparison between the predicted and experimentally-derived values was performed using mean squared error and the Pearson sample correlation coefficient.

Key findings The findings of this study would suggest that the models failed to accurately predict permeation and in some cases were not within two- or three-orders of magnitude of the experimentally-derived values. However, with this set of compounds the models were able to effectively rank the permeants.

Conclusions Although not suitable for accurately predicting permeation the models may be suitable for determining a rank order of permeation, which may help to select candidate molecules for in-vitro screening. However, it is important to note that such predictions need to take into account actual relative drug candidate potencies.

Introduction

The developments in combinatorial chemistry and high throughput screening have led to a rapid increase in drug design and discovery and, ultimately, the production of many potential molecules that require evaluation. As such, the use of empirical or traditional screening methods such as *in-vivo* and in-vitro testing is often restricted due to the time and cost involved. Thus, modern drug selection requires rapid and cost effective methods that are applicable to a large number of samples, one example being the use of mathematical models. Hence, over the last 25 years there has been much interest in the use of such models to predict dermal absorption *in silico*.

It is well established that the physicochemical properties of a molecule exert a substantial effect on its permeability, and as such formulation scientists have been tempted, especially in the pharmaceutical arena, to find the most promising candidate molecules by using a mathematical relationship between percutaneous permeation and molecular parameters. Generally this involves the use of the discrete descriptors of a molecule, such as lipophilicity (most commonly expressed as log P, the logarithm of the octanol-water partition coefficient), hydrogen bonding, molecular weight (or size), melting point and solubility parameter, often in the form of a quantitative structure permeability relationship (QSPR).[1–10]

For many QSPRs, regression analysis is the statistical method of choice.[11–13] However, there are a number of disadvantages to using this method, which have been addressed in detail previously.[14] Firstly, it is a linear technique, and secondly, it may be adversely affected by any co-linearity

between seemingly independent variables such as log P and molecular weight. Despite the large body of literature that exists in this field, recent research has suggested that methods based on linear regression analysis may not be an entirely suitable tool for the development of QSPRs; nor is it clear whether linearity is appropriate for the modelling of highly hydrophilic and hydrophobic molecules. For example, Moss *et al.*[8] compared the statistical accuracy of Gaussian processes (GP), single linear networks and QSPRs by a range of statistical methods, and found that the underlying nature of the dataset was inherently nonlinear (i.e. the data were modelled by a function which was a nonlinear combination of the model parameters and was dependent upon one or more variables) and that skin permeation (as represented by K_p, which describes the rate of permeant transport per unit concentration, given in units of distance/time (e.g. cm/h)) was best described, in purely statistical terms, by GP approaches.

A range of nonlinear methods have been employed to investigate predictions of skin absorption. Artificial neural networks have been investigated and showed good predictive ability.[15] However, artificial neural networks are a somewhat limited method in that they have a tendency to over-fit where large numbers of physicochemical descriptors exist, compared with the data points used. Such models are often weighted and are susceptible to over-training.[16] GP methods do not alleviate all these issues, but minimise them, reportedly providing better predictions of percutaneous absorption than existing models.[8,10,17,18] Moreover, Lam *et al.*[10] demonstrated that a certain degree of inter-changeability existed between physicochemical parameters used for developing models of percutaneous absorption, and that the development of models which explicitly represent their output as a defined equation, with discrete parameters, may not, from a mechanistic point of view, entirely represent the underlying nature of the skin absorption process.

In addition, the datasets used to develop the models described above have been derived from experimentally-derived permeation data.[19–23] The quality of such experimentally-derived data relies upon the quality of the experiments performed and some form of standardization of the methodologies. A clear example of this is the difference between the Potts and Guy[3,21] models, which employed different datasets (the latter model examining a subset of the former model's dataset) and yielded two models that provide two very different representations of skin absorption.

Therefore, the aim of this study was to test the performance of both linear and nonlinear models to predict the skin permeation of a series of 11 candidate test compounds belonging to the same therapeutic class. As these compounds were chosen from different chemical classes they represented a wide range of physicochemical parameters but were within the scope of the models, and the datasets which were used to derive them.

Materials and Methods

Materials

Ammonium hydrogen carbonate was purchased from Sigma-Aldrich (Gillingham, Dorset, UK). Disodium hydrogen phosphate was supplied by Fluka (Gillingham, Dorset, UK). Acetic acid, glycerol and polyethylene (PEG) 400 were obtained from Merck (Merck Chemicals Ltd., Nottingham, UK). Brij 98 was provided by Croda (Croda Europe Ltd., Snaith Goole, East Yorkshire, UK). Phosphate buffered saline (PBS) was purchased from Oxoid (Basingstoke, Hampshire, UK). Citric acid anhydrous was purchased from VWR (VWR International, Lutterworth, Leicestershire, UK). High performance liquid chromatography (HPLC) grade acetonitrile was purchased from Fisher (Fisher Scientific, Loughborough, Leicestershire, UK). Human skin was donated from a healthy volunteer following approval by the Schools of Pharmacy and Postgraduate Medicine Ethics Committee with Delegated Authority (Study Ref. No. PHAEC/09–23). Validated Franz permeation cells were donated by MedPharm Ltd. (Guildford, UK). Deionised water was obtained from a Millipore Water Purification system (MilliQ, 18.2 MΩ).

All of the test compounds utilised in the skin permeation studies were provided by Nycomed (Konstanz, Germany). Their physicochemical properties are summarised in Table 1.

Mathematical prediction of skin absorption

The modelling in this study was carried out by the application of a combination of QSPRs and GP-based machine learning methods.[3,8,10,24]

Potts and Guy (1992) model

The Potts and Guy (1992) equation is summarised in equation 1 for the calculation of the permeation coefficient across the epidermal membrane:

$$K_p(cm/h) = 3600\ anti\log(-6.3 + 0.71\log K_{o/w} - 0.0061\,MW) \quad (1)$$

Where K_p is the permeation coefficient of the drug across the epidermal membrane (eqn 1), MW is the molecular weight of the compound, and log $K_{o/w}$ is the octanol/water partition coefficient.[3]

Equation 1 was employed for the Potts and Guy model in this study so that comparison could be made with the K_p value calculated using other methods, discussed below.

Wilschut–Robinson model

The Wilschut–Robinson model is shown in equation 2.[24] This model considers the heterogeneity of the skin by calculation of different permeation coefficients for the lipid fraction ($K_{p,sc}$; eqn 3) and the protein fraction (K_{pol}; eqn 4) of the stratum corneum, and through the hydrophilic epidermal

Table 1 Physicochemical properties of the 11 test compounds tested for skin permeation

Compound number	Aqueous solubility (mg/ml)	Solubility in donor fluid (mg/g)	Log P	Molecular weight (g/mol)	Solubility parameters by Fedors	Number of hydrogen bond acceptors (HBA)	Number of hydrogen bond donors (HBD)
1	0.00062	6.8	3.96	403.22	14.0739	7	1
2	0.57747	25.88	2.65	412.48	12.2706	7	1
3	0.25879	0.215	0.94	470.52	15.9180	10	0
4	1.41877	15.955	1.82	383.45	12.8590	7	1
5	0.03606	2.68	3.00	508.57	15.7335	10	0
6	0.41668	14.49	1.93	462.98	13.6399	8	2
7	0.00093	1.2	4.09	489.58	11.2123	6	0
8	0.00039	0.07	3.02	436.52	11.9148	5	1
9	0.00298	0.245	3.50	465.55	11.5695	6	1
10	0.01048	15.93	4.80	436.62	9.9744	3	0
11	0.00001	0.805	5.13	817.01	12.6723	9	0

layer (K_{aq}; eqn 5), before calculation of the overall permeation coefficient (K_p; eqn 2).

Overall permeation coefficient:

$$K_p\,(cm/h) = \frac{1}{\dfrac{1}{K_{p,sc} + K_{pol}} + \dfrac{1}{K_{aq}}} \tag{2}$$

Where K_p is the overall permeation coefficient (eqn 2), $K_{p,sc}$ is the permeation coefficient of the compound through the lipid fraction of the stratum corneum (eqn 3), K_{pol} is the permeation coefficient of the compound through the protein fraction of the stratum corneum (eqn 4), and K_{aq} is the permeation coefficient of the compound through the watery epidermal layer (eqn 5).

The permeation coefficient of the compound through the lipid fraction of stratum corneum is described by equation 3:

$$\log K_{p,sc} = -1.326 + 0.6097x \log K_{o/w} - 0.1786x\, MW^{0.5} \tag{3}$$

The permeation coefficient of the compound through the protein fraction of stratum corneum is described by equation 4:

$$K_{pol} = \frac{0.0001519}{\sqrt{MW}} \tag{4}$$

The permeation coefficient of the compound through the watery epidermal layer is described by equation 5:

$$K_{aq} = \frac{2.5}{\sqrt{MW}} \tag{5}$$

Moss and Cronin[14] quantitative structure-permeability relationship model

The QSPR by Moss and Cronin[14] was employed in an attempt to correlate the structural or property descriptors of the compounds with the prediction of skin permeation. This model is summarised in equation 6:

$$K_p\,(cm/h) = anti\log\,(-2.55 - 0.00389\,MW + 0.357\log K_{o/w} - 0.02\,HBA - 0.107\,HBD) \tag{6}$$

Where K_p is the overall permeation coefficient (eqn 6), HBA is the hydrogen bond acceptor group, HBD is the hydrogen bond donor group, *MW* is the molecular weight of the compound, and log $K_{o/w}$ is the octanol/water partition coefficient.

5f model

The final QSPR-type model employed in the calculation of the permeation coefficient (K_p) was the 5f model (eqn 7).[25] This model incorporates the solubility parameters (*SP*) as predicted using Fedors solubility parameter.[26]

$$K_p\,(cm/h) = anti\log\,(-2.236 - 0.00388\,MW - 0.01367\,SP + 0.2634\log K_{o.w} - 0.0787\,HBA - 0.01136\,HBD) \tag{7}$$

Where K_p is the overall permeation coefficient (eqn 7), *HBA* is the number of hydrogen bond acceptor groups, *HBD* is the number of hydrogen bond donor groups, *MW* is the molecular weight of the compound, log $K_{o/w}$ is the octanol/water partition coefficient, and *SP* is the solubility parameter calculated using Fedors' model.[26]

Gaussian process regression

GPs are non-parametric methods of modelling. They do not output a specific functional representation of the data, as the methods used to generate QSPR models do, in the form of an explicit mathematical relationship with discrete, statistically significant parameters. In GP modelling it is assumed that the underlying function that produces the data, $f(x)$, remains

unknown, but that the data are produced from a (infinite) set of functions, with a Gaussian distribution in the function space.[8,17]. A GP is characterised fully by its mean and the covariance function. The mean is normally considered to be the 'zero everywhere' function, and the covariance function, $k(x_i, x_j)$, expresses the expected correlation between the values of $f(x)$ at the two points x_i, x_j, defining nearness, or similarity, between data points. The mean at x_* is given by equation 8:

$$E\left[y_*\right] = k_*^T \left(K + \sigma_n^2 I\right)^{-1} y \quad (8)$$

Where k_* denotes the vector of covariances between the test point and N_{trn} training data, K denotes the covariance matrix of the training data, σ_n^2 is the variance of an independent identically distributed Gaussian noise, which means that observations are noisy, k_*^T is the transpose of k_*, and I is the identity matrix, finally, y denotes the vector of training targets. The variance, at x_*, is given by equation 9:

$$\mathrm{var}\left[y_*\right] = k(x_*, x_*) - k_*^T \left(K + \sigma_n^2 I\right)^{-1} k_* \quad (9)$$

Where $k(x_*, x_*)$ is the variance of y_*.

In this study, the mean has been used as the prediction and the variance as error bars on the prediction.

Rate of permeation (flux) calculations

The predicted rate of permeation (maximum flux or *Max J*) of the compounds through the skin (flux, mg/(cm^2.h)) was calculated, using equation 10:

$$Max\ J = K_p\ x\ S \quad (10)$$

Where *Max J* is the maximum flux (in mg/(cm^2.h)), S is the solubility (in mg/ml), and K_p is the permeation coefficient (in cm/h).

Human skin preparation

Human skin was sourced from one single donor with full ethics permission (BMSEP08/05/04E2, female Asian aged 60) via a cosmetic reduction surgery (taken from an abdominoplasty) and was prepared using the method described by Kligman and Christophers.[27] Full thickness skin previously frozen (–20°C) was defrosted at ambient temperature until malleable, at which time the subcutaneous fat was removed mechanically by blunt dissection. The skin was then immersed in deionised water heated to 60°C for 45 s. The epidermal membrane (comprising the stratum corneum and epidermis) was removed from the underlying dermis using a gloved finger and the dermis discarded and then floated (stratum corneum side up) in deionised water and onto filter paper. The epidermal membrane mounted on filter paper was removed from the deionised water, any excess water was removed using tissue and the epidermal membrane was either used immediately or stored at –20°C until use.

Test compound in-vitro skin permeation studies

Static vertical Franz type diffusion cells (MedPharm Ltd.) with a diffusion surface area of approximately 0.6 cm^2 were individually calibrated. The epidermal membrane was sandwiched between the donor and receiver chamber of the diffusion cells, which were then sealed with Parafilm. The receiver chamber was then filled with the receiver fluid (pH 5 phosphate-citrate buffer containing 20% PEG 400 and 2% Brij 98) to ensure sink conditions were ensured. A small PTFE-coated magnetic stir bar was placed inside the receiver chamber through the sampling arm and driven by a motorless stirrer plate. After cell equilibration at 37°C for one hour, infinite doses (100 µl) of saturated solutions (PEG 400 : glycerol (50 : 50)) of all 11 candidate molecules were applied to the surface of each epidermal membrane ($n = 6$ per candidate molecule). At predetermined time intervals (1, 3, 5, 10, 22, 26, 30, 34 and 48 h), 100 µl was removed via the sampling arm from the donor chamber points with a 100-µl Hamilton syringe and the candidate concentration measured by HPLC. The sample volume was replaced by an equivalent volume of fresh receiver fluid to maintain a constant volume. The permeation studies lasted a period of six weeks.

High performance liquid chromatography analysis

Samples were stored at 2–8°C before analysis using a Waters 2695 Alliance HPLC system with Waters 996 Photo-diode array detector using Waters Empower^2 Data Processing. The column used was a Phenomenex Gemini C18, 50 × 4.6 mm, 5 µm, 110 Å fitted with a SecurityGuard Analytical Cartridge Holder (4 mm) and Gemini Security guard cartridges (C18, 4 × 3.0 mm) maintained at 50 ± 2°C with an injection volume of 10 µl. For all candidate molecules a gradient method was used comprising mobile phase A (ammonium hydrogen carbonate (20 mM, pH 7.8)) and mobile phase B (100% acetonitrile).

For test compounds 1, 7, 10 and 11 the gradient was 0–30 min 90–20% mobile phase A; 30–32 min 20–90% mobile phase A, which was then maintained for a further 8 min. The detection wavelengths for the four candidates were 248, 300, 257 and 316 nm, respectively. For test compounds 2, 3, 4, 5, 6, 8 and 9 the gradient was 0–25 min 90–31% mobile phase A; 25–26 min 31–90% mobile phase A, which was then maintained for a further 4 min. The detection wavelengths for the seven candidates were 230, 316, 235, 317, 317, 235 and 316 nm, respectively.

Data analysis

The cumulative amounts of the candidate molecule (μg) penetrating the unit surface area of skin (cm^2) were corrected for sample removal and plotted against time (h). Steady-state flux was calculated using the linear regression of the cumulative amount per unit area ($\mu g/cm^2$) against time comprising five time points from $t = 22$–48 h for each compound (linearity of $r^2 \geq 0.98$) using Microsoft Office Excel 2003.

K_p and flux calculations

For the mathematical predictions, K_p is predicted and thus the predicted flux was calculated using either equation 11 or 12 based on aqueous solubility and mean saturated solubility in the donor fluid, respectively:

$$Predicted\ flux = Predicted\ K_p \times aqueous\ solubility \quad (11)$$

$$Predicted\ flux = Predicted\ K_p \times mean\ of\ saturated\ solubility\ in\ donor\ fluid \quad (12)$$

The experimental K_p was calculated using equation 13:

$$Experimental\ K_p = \frac{Experimental\ flux}{Mean\ of\ saturated\ solubility\ in\ donor\ fluid} \quad (13)$$

Performance measurements (prediction to actual)

A number of measures of the statistical performance of the GP models were employed.[8,10] Mean squared error (MSE) measures the average squared difference between model predictions y_i (theoretically predicted data) corresponding targets x_i (experimental data). It is defined in equation 14:

$$MSE = \frac{1}{N}\sum_{n=1}^{N}(x_i - y_i)^2 \quad (14)$$

Where N is the number of test data points.

The Pearson sample correlation coefficient, r, between targets and predictions is employed to assess the extent of a linear relationship between two datasets of a model. It is defined by dividing the covariance of the two variables (targets x_i and model predictions y_i) by the product of their standard deviations, shown in equation 15:

$$r = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \quad (15)$$

Where cov is the covariance of the two variables, and σ is the standard deviation of each variable.

When results are analysed, a model that provides a low value on MSE and a high value on the Pearson sample correlation coefficient r is preferred.

Results

Predicted and actual correlation

Figure 1 depicts the amount of each test compound permeating across epidermal membrane over 48 h. Table 2 provides the results of the modelling data for the prediction of K_p and compares it with the K_p measurements experimentally derived. Figure 2 provides a diagrammatic representation of a comparison of the K_p values between the experimentally derived data and that predicted from the five models. Test compound 8 was observed not to permeate within $t = 48$ h and therefore it has not been included. The data suggested, albeit in a qualitative sense, that the 5f model was the best at predicting K_p as for nearly all the candidates this model was closest to the experimental data and this was especially true for candidates 5, 6 and 9 where there was almost complete overlap. For the remainder of the candidates the models generally over-predicted the K_p values, as demonstrated by the fact that the experimentally-derived data was at the core of Figure 2.

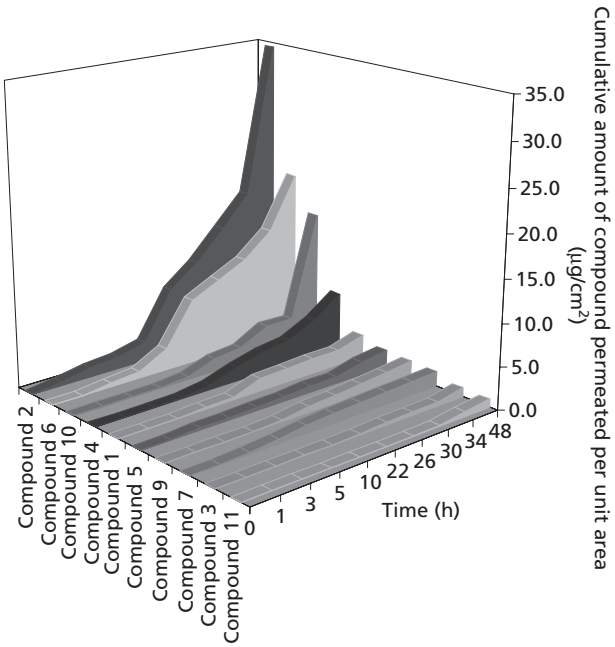

Figure 1 The mean cumulative amount of the test compounds permeated per unit area against time across human epidermal membrane. The data is presented as the mean of $n = 3$–6.

Table 2 The values of the experimental and the predicted permeation coefficient of the drug across the epidermal membrane (K_p)

Test compound	Experimental K_p (cm/h)	Prediction of K_p (cm/h) based on a specific mathematical model				
		Potts and Guy	Robinson	Quantitative structure permeability relationships	5f model	Gaussian process
1	7.00E-06	4.06E-03	3.11E-03	1.12E-03	3.07E-04	7.02E-04
2	4.05E-05	4.18E-04	4.65E-04	3.50E-04	1.35E-04	7.79E-04
3	1.58E-04	1.13E-05	3.06E-05	5.69E-05	1.52E-05	6.37E-04
4	1.26E-05	1.62E-04	2.01E-04	2.30E-04	1.04E-04	6.71E-04
5	2.14E-05	1.92E-04	3.05E-04	2.20E-04	3.79E-05	6.37E-04
6	3.23E-05	6.34E-05	1.09E-04	9.20E-05	4.34E-05	4.99E-04
7	1.85E-05	1.49E-03	1.63E-03	7.69E-04	2.07E-04	7.07E-04
8	0.00E+00	5.47E-04	6.12E-04	4.20E-04	1.99E-04	1.10E-03
9	1.88E-04	7.97E-04	9.00E-04	4.59E-04	1.73E-04	8.68E-04
10	3.49E-05	1.00E-02	6.97E-03	2.54E-03	9.15E-04	1.08E-03
11	4.87E-05	8.22E-05	5.00E-04	8.38E-05	1.16E-05	6.33E-04

The experimental *Kp* was calculated from the division of experimental flux provided in 4 by the compound solubility in the donor fluid. Table 4 also provides details of the levels of experimental variability. Please note that the test compound 8 was observed not to permeate within $t = 48$ h permeation experiment, therefore the experimental *Kp* is zero.

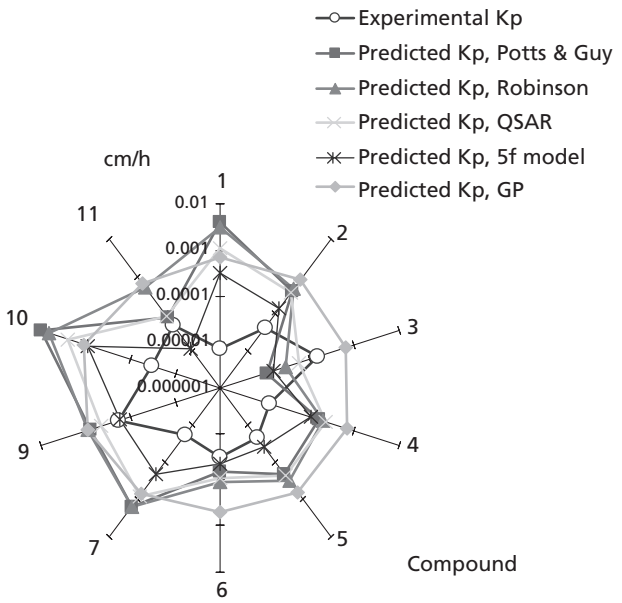

Figure 2 Comparison of experimental permeation coefficient of the drug across the epidermal membrane (K_p) to predicted K_p using five mathematical models. The experimental K_p was calculated from the division of experimental flux by the compound solubility in donor fluid. Please note that test compound 8 was observed not to permeate within $t = 48$ h permeation experiment, therefore it has not been included.

This observation was confirmed statistically by the data in Table 3, which shows the MSE results where the 5f model was observed to outperform all the other models, whilst GP gave the poorest correlation on this test set. However, Table 3 shows the Pearson correlation results obtained for the data, indicating that the degree of linear dependence between experimental (target) values and predictions was poor amongst all models, with again the 5f model giving the best positive linear relationship to the experimental values, followed by the GP model. Moreover, it could be seen that the Potts and Guy, Robinson and QSPR models had negative linear relationships to the target values, with the QSPR model giving the strongest negative linear dependence.

Table 4 provides the results of the modelling data for the prediction of flux using both aqueous solubility and donor fluid solubility. Figure 3 provides a diagrammatic representation of a comparison of the flux values between the experimentally-derived data and that predicted from the five models. The data suggested that, when aqueous solubility was used to calculate flux, all models under-predicted the measured flux values with no model showing an obvious benefit. This was the opposite of what was observed when the donor fluid solubility was used and the models over-predicted the flux values, with the 5f model being the best performer.

However, the data presented in Tables 5 and 6 provide a statistical evaluation of the data and suggested that when aqueous solubility was used (Table 5) to determine the flux then the Potts and Guy (1992) model followed by the Robinson model were the most accurate in terms of the Pearson correlation with the order reversed for the MSE. the Moss QSPR model was third in both; with all 3 models being linear in nature. However, when donor fluid solubility was used (Table 6) the GP model was the most accurate, in fact this gave the best correlation of all the models and solubilities when using Pearson's correlation followed by the 5f model, but the order of these two were reversed for the MSE. In both calculations the QSPR, Robinson and Potts and Guy were the worst performing, in descending order.

Table 3 Pearson correlation coefficients and mean squared error between prediction of K_p (the permeation coefficient of the drug across the epidermal membrane) based on aqueous solubility and their corresponding experimental values for test compounds 1–11

	Comparison between prediction of K_p and their corresponding experimental values				
Parameter to compare	Potts and Guy	Robinson	Quantitative structure permeability relationships	5f model	Gaussian process
Pearson correlation	–0.1723	–0.1804	–0.1965	0.1708	0.0895
Mean squared error	1.1023E-005	5.6113E-006	7.7499E-007	8.9030E-008	1.4239E-004

Ranking comparison

The data detailed in Tables 2 and 4 were used to rank the candidate compounds in terms of K_p and flux in Tables 7 and 8, respectively. For K_p the rank orders produced in the models showed little comparison with the experimentally determined data. The top ranking candidate for K_p from the models was compound 10, which ranked fifth in the experimentally-derived data. All four of the linear models had candidates 10, 1, 7, 9 and 8 in their top five for predicted K_p , but only compounds 10 and 9 where in the top five for the experimental data. The nonlinear GP model ranked three (i.e. 2, 9, 10) molecules that were in the top five determined experimentally.

The predicted flux data calculated using aqueous solubility gave a much closer similarity in ranking with the models. The Potts and Guy, and Robinson models predicted all of the top five molecules determined experimentally with the Moss QSPR, 5f and GP predicting four out of five. The ranking predictions were less reliable when using the donor fluid solubility to predict flux with the Moss QSPR and 5f model predicting four and the Potts and Guy, Robinson and GP models predicting three of the top five molecules using the experimentally-determined flux values.

Discussion

The findings of this study indicated a poor correlation between a range of mathematical models and experimentally measured permeability – of different types – for prediction of percutaneous absorption. While experimental data has been shown to possibly vary between experiments this study has focussed on results from one laboratory and used skin donated from one volunteer.[28] The experimental results also showed a similar level of variability compared with any number of similar studies in the literature. Therefore, while skin from only one donor has been used there was little to suggest that skin from this particular donor was an 'outlier' and the results were consistent with those expected in the literature from compounds with similar physicochemical properties. It may also be suggested that using skin from one donor has, in an experiment of this nature, a significant advantage in that it may reduce variability and allow a reasonable comparative estimate of permeation to be made. From a practical perspective human skin is a scarce commodity and it is often not possible to perform such studies with skin from multiple skin donors making the study a more realistic comparison with that performed industrially. Therefore, the key finding of this study was the failure of a wide range of molecular models for percutaneous absorption to accurately predict the permeability of a range of compounds that varied widely in their physicochemical properties. Further, the models each offered significantly different predictions of permeability; while the predictions from the QSPR models appeared to be well correlated with each other they showed no commonality with the GP model.

The data used in this study, particularly the 11 test compounds, was presented in a similar manner to many similar studies published. Specifically, the aim of this study was to examine the use and accuracy of models, rather than to add to or develop those models. Addition of the chemical structures, to a group of chemicals that fit the chemical space (i.e. the test chemicals did not exceed the boundaries of the models) described by the dataset used for model construction, added no additional useful information, particularly as the important physicochemical parameters used to produce predictions has been provided in Table 1, as well as information on penetrant solubility. Further, as Moss *et al*.[29] have shown, there is little to be gained from adding more and more data to the same 'chemical space' in a model, and the possibility of skewing the data may indeed raise issues regarding the quality of dataset construction that would not be aided by adding more – and similar – chemicals to the dataset. Ultimately, it may be suggested that, as the aim of this study was to compare the accuracy of model predictions with experimental results, the absence of chemical structures in such a circumstance neither adds nor detracts from this work.

When flux was assessed using the aqueous solubility of the penetrants it was under-predicted, relative to experimental measurements, and over-predicted when the solubility of the penetrant in the donor phase solution was used. It also appeared that there was no obvious benefit in using experimentally-derived donor solubility rather than the experimentally-measured aqueous solubility in these calculations. In addition, while the absolute correlations were often poor (in some cases, differing from the experimental data by

Table 4 The values of the experimental flux and the predicted flux (μg/(cm².h) based on the reported aqueous solubility and solubility in donor fluid

Compound	Experimental flux (μg/(cm².h)) Mean ± SEM (n = 3–6)	Prediction of flux (μg/(cm².h)) based on aqueous solubility				Prediction of flux (μg/(cm².h)) based on solubility in donor fluid				
		Predicted flux, Potts & Guy (aq sol)	Predicted flux, Robinson (aq sol)	Predicted flux, QSPR (aq sol)	Predicted flux, 5f model (aq sol)	Predicted flux, GP (aq sol)	Predicted flux, Potts & Guy (donor sol)	Predicted flux, Robinson (donor sol)	Predicted flux, QSPR (donor sol)	Predicted flux, 5f model (donor sol)	Predicted flux, GP (donor sol)
1	0.0476 ± 0.0156	2.52E-03	1.93E-03	6.94E-04	1.91E-04	4.36E-04	2.76E+01	2.11E+01	7.60E+00	2.09E+00	4.77E+00
2	1.0481 ± 0.2106	2.42E-01	2.69E-01	2.02E-01	7.81E-02	4.50E-01	1.08E+01	1.20E+01	9.07E+00	3.50E+00	2.02E+01
3	0.0339 ± 0.0151	2.93E-03	7.92E-03	1.47E-02	3.93E-03	1.65E-01	2.43E-03	6.58E-03	1.22E-02	3.26E-03	1.37E-01
4	0.2007 ± 0.0505	2.30E-01	2.85E-01	3.26E-01	1.48E-01	9.52E-01	2.58E+00	3.20E+00	3.66E+00	1.66E+00	1.07E+01
5	0.0573 ± 0.0278	6.93E-03	1.10E-02	7.94E-03	1.37E-03	2.30E-02	5.15E-01	8.17E-01	5.90E-01	1.02E-01	1.71E+00
6	0.4685 ± 0.2833	2.64E-02	4.53E-02	3.84E-02	1.81E-02	2.08E-01	9.19E-01	1.58E+00	1.33E+00	6.28E-01	7.23E+00
7	0.0222 ± 0.0050	1.39E-03	1.51E-03	7.15E-04	1.93E-04	6.58E-04	1.79E+00	1.95E+00	9.22E-01	2.48E-01	8.49E-01
8	0.0000 ± 0.0000	2.15E-04	2.40E-04	1.65E-04	7.80E-05	4.32E-04	3.83E-02	4.28E-02	2.94E-02	1.39E-02	7.70E-02
9	0.0461 ± 0.0186	2.37E-03	2.68E-03	1.37E-03	5.15E-04	2.59E-03	1.95E-01	2.21E-01	1.12E-01	4.23E-02	2.13E-01
10	0.5557 ± 0.0728	1.05E-01	7.30E-02	2.66E-02	9.59E-03	1.13E-02	1.60E+02	1.11E+02	4.05E+01	1.46E+01	1.72E+01
11	0.0392 ± 0.0115	5.37E-07	3.27E-06	5.48E-07	7.57E-08	4.14E-06	6.62E-02	4.02E-01	6.75E-02	9.32E-03	5.10E-01

Aqueous solubility, aq sol; solubility in donor fluid, donor sol.

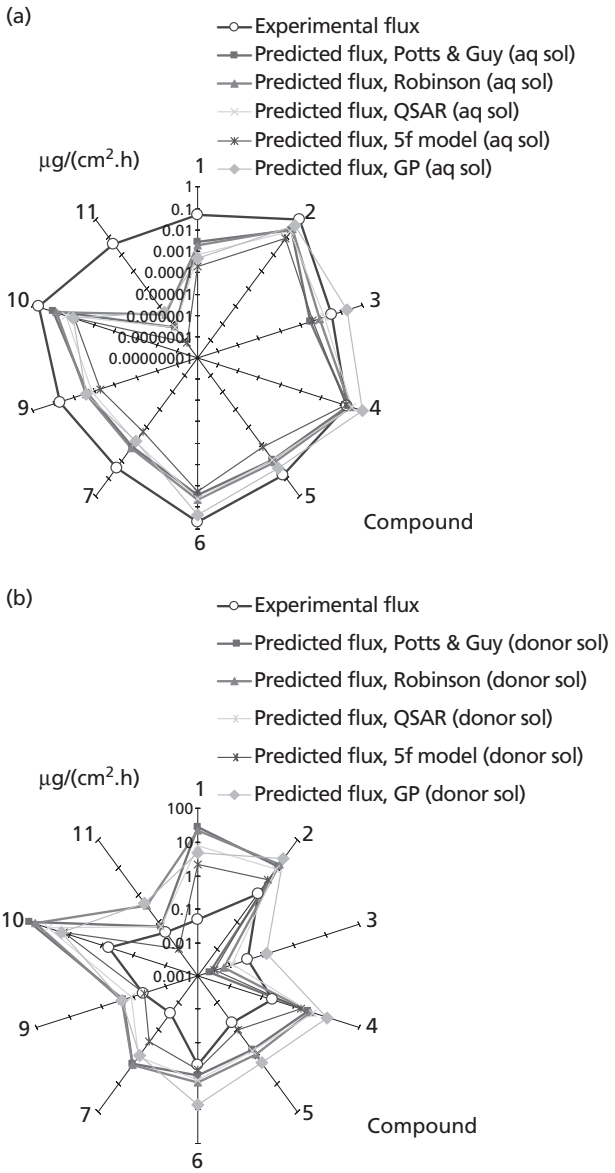

Figure 3 Comparison between the prediction of flux based on aqueous solubility and the solubility in donor fluid, with the experimental flux for the test compounds. (a) Aqueous solubility, aq sol; (b) donor fluid, donor sol. Please note that test compound 8 was observed not to permeate within t = 48 h permeation experiment, therefore it has not been included.

three- or four-orders of magnitude) the rank order of permeability (Table 8) was relatively consistent with calculation of flux using experimental aqueous solubility giving a much closer rank order than that when the donor fluid solubility was used.

In Cronin and Schultz's[11] review of the pitfalls of QSPR models, they listed a series of criteria for the use and analysis of those models. Among their key recommendations were the

Table 5 Pearson correlation coefficients and mean squared error between prediction of flux based on aqueous solubility and their corresponding experimental values for the test compounds 1–11

	Comparison between prediction of flux based on aqueous solubility and their corresponding experimental values				
Parameter	Potts and Guy	Robinson	QSPR	5f model	GP
Pearson correlation	0.7354	0.6754	0.4763	0.4216	0.3637
Mean squared error	0.0962	0.0941	0.1096	0.1323	0.1192

Table 6 Pearson correlation coefficients and mean squared error between prediction of flux based on solubility in donor fluid and their corresponding experimental values for the test compounds 1–11

	Comparison between prediction of flux based on solubility in donor fluid and their corresponding experimental values				
Parameter	Potts and Guy	Robinson	QSPR	5f model	GP
Pearson correlation	0.3527	0.3837	0.4837	0.5095	0.9142
Mean squared error	2.3896E+003	1.1614E+003	157.3405	19.0578	75.0694

Table 7 Test compound ranking based on the experimental and the predicted permeation coefficient of the drug across the epidermal membrane$_p$)

		Compound ranking based on the predicted K_p				
Ranking	Test Compound ranking based on experimental K_p	Potts and Guy	Robinson	QSPR	5f model	GP
1	9	10	10	10	10	8
2	3	1	1	1	1	10
3	11	7	7	7	7	9
4	2	9	9	9	8	2
5	10	8	8	8	9	7
6	6	2	11	2	2	1
7	5	5	2	4	4	4
8	7	4	5	5	6	3
9	4	11	4	6	5	5
10	1	6	6	11	3	11
11	8	3	3	3	11	6

The permeation coefficient of drug across the epidermal membrane, K_p.

avoidance of extrapolation beyond the original domain of the QSPR, appreciation of the precision of the QSPR and its expected application in the context of the original biological measurement (in the case of percutaneous absorption, J_{max} or K_p), a single model describing the whole process, not one that describes single steps or parts of the dataset, the avoidance of nontransparent QSPRs (or, in the context of this study, the avoidance of an approach such as the 'black-box' nature of GP models) and the correct understanding of the endpoint of the QSPR and its intended scope of use. Finally, they recommended the development of QSPRs by multi-disciplinary groups of researchers whose expertise extends across all parts of the study and its methodology. Their latter point was particularly well emphasised in this, and previous studies.[8,10] It is also important to build upon Cronin and Schultz's comment on the fundamental nature of a QSPR or other 'mathematical model' of a biological process, which they defined as a triangulation of biological endpoint data, physicochemical and structural information of the chemicals of which the dataset is comprised and a suitable mathematical and/or statistical approach to model development. This suggested that, rather than considering simply the physicochemical properties of a penetrant, it may be important to consider and include biological data in the predictions.

Another key issue in the construction of a multidisciplinary team to examine this problem 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 nonlinearity of the dataset used for percutaneous absorption modelling, and highlighted the issue with many studies that have adopted linear methods in their analysis.[8,14,18] While these previous studies have been benchmarked against the 1992 model of Potts and Guy[3], due to its enormous contribution to this field and its widespread acceptance as the first quantitative model of percutaneous

Table 8 Test compound ranking based on experimental and predicted flux values derived from the reported aqueous solubility and solubility in donor fluid

		Ranking based on predicted flux derived from aqueous solubility					Ranking based on predicted flux derived from solubility in donor fluid				
Ranking	Ranking based on experimental flux	Potts and Guy (aq sol)	Robinson (aq sol)	QSPR (aq sol)	5f model (aq sol)	GP (aq sol)	Potts and Guy (donor sol)	Robinson (donor sol)	QSPR (donor sol)	5f model (donor sol)	GP (donor sol)
1	2	2	4	4	4	4	10	10	10	10	2
2	10	4	2	2	2	2	1	1	2	2	10
3	6	10	10	6	6	6	2	2	1	1	4
4	4	6	6	10	10	3	4	4	4	4	6
5	5	5	5	3	3	5	7	7	6	6	1
6	1	3	3	5	5	10	6	6	7	7	5
7	9	1	9	9	9	9	5	5	5	5	7
8	11	9	1	7	7	7	9	11	9	9	11
9	3	7	7	1	1	1	11	9	11	8	9
10	7	8	8	8	8	8	8	8	8	11	3
11	8	11	11	11	11	11	3	3	3	3	8

Aqueous solubility, aq sol; solubility in donor fluid, donor sol.

absorption, iterations of that model which include non-linear terms might provide more substantial and realistic benchmarks.

The reasons for the discrepancies between the modelled and measured permeation values for each of the 11 compounds evaluated may have been, in part, due to the experimental protocol used including the type of donor and receiver systems, type of diffusion cell, skin source etc. However, the methodology used was a realistic and practical representation of what would be performed from day to day in the pharma industry. In addition, it was apparent that the 11 test chemicals covered a wide range of 'chemical space'. Therefore, as has been discussed previously, the nature of the model, particularly if it is based on a linear regression method, may produce an inaccurate prediction of permeability.[8,10,18] This was shown when the predictively of the Potts and Guy[3] equation was compared with the GP model.[9,10] Therefore, the differences in prediction may have been in part due to the inability of certain models to accurately predict permeability across a wide range of chemical space, an issue which the GP methods appeared better positioned to deal with due to their inherent nonlinear nature as well as the size and scope of the dataset used to construct them.[9]

Extrapolation of a model beyond its boundaries is not common sense and a clear limitation to 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, it might also be difficult to develop a model that is uniformly accurate even within these boundaries. Sun *et al.*[18] demonstrated the substantial increases in covariance as the GP models extended outside those areas of the dataset that were highly populated. While this also implied criticism of comparisons made between GP models and, for example, the Potts and Guy[3] equation, it is important to demonstrate the fallibility of models in such circumstances to avoid inappropriate use. This was highlighted by Moss *et al.*[30], who showed how QSPR models of percutaneous absorption failed to accurately predict the permeability coefficient, K_p, across a range of log *P*-values. While those authors highlighted that point, they also clearly demonstrated the general failure of prediction across the whole range under examination. In this study the failure of models to predict a wide range of permeability (in terms of physicochemical properties) suggested that the models developed were quite poor in terms of their generic applicability. It should also be noted that the models developed previously (i.e. Moss *et al.* [8] and Sun *et al.*) were based on datasets that were approximately 50% larger than those of, for example, Potts and Guy.[3] Hence, while modelling outside the ranges of a dataset is not desirable, the GP method appeared to be able to analyse such data points better than linear models. Such discussions inevitably lead onto the subject of outliers which have been described elsewhere.[14]

One point not directly raised by Cronin and Schultz[11] was the quality of the data used to construct the model. 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, however, it is critical to note that the quality of any model is completely dependent upon the quality of the data from which it was derived.[14] A superficial review of the skin absorption literature quickly highlights the lack of uniformity in skin permeation techniques used between well established and respected groups, raising the question again of why there is still no standardised protocols rather than guidelines available for skin permeation experiments. In addition, 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[31] analysed a subset of the Flynn[19] dataset which comprised of 22 compounds. Other studies have, for various reasons (i.e. Potts and Guy[21]) examined only nonelectrolytes from the Flynn dataset, selected subsets of a larger database or examined a small number of compounds and derived QSPR-type models. These range from the work of Barratt[20], which examined 60 'small molecules and steroids' excluding the hydrocortisone derivatives, from the Flynn dataset, Abraham *et al*.[6,32] who examined, respectively, 46 and 53 compounds, to those who analysed substantially smaller datasets with 20, 16 (in different studies by both Lee *et al*.[34] and Morimoto *et al*.[35]) or four compounds.[33,36] In all those cases, mechanistic inferences were drawn into the percutaneous absorption of the compounds and, by inference, those that were chemically similar. However, given the issues presented in this study the number of compounds present in those datasets would suggest that the value of those models may be limited by the amount and quality of available data.

None of the mathematical models used to predict *Kp* take into account the ionisation state of a compound, which can ultimately affect its partitioning and diffusion into and across the stratum corneum. However, it is also important to note that although aqueous solubility was used to calculate flux from the predicted K_p values, during the in-vitro permeation studies, a nonaqueous system comprising of a mixture of PEG 400 and glycerol was employed owing to the fact that some compounds were observed to be unstable in an aqueous system. In addition, a higher extent of solubility could be achieved in this nonaqueous system for the purpose of in-vitro assessment, in contrast to a lower achievable concentration in an aqueous system. Therefore, it was assumed that during the in-vitro permeation across human epidermal membrane, there was no ionisation for those ionisable compounds (acidic or basic) in this nonaqueous system. Thus the lack of consideration of the ionisation state could not explain the poor correlations observed, although this does warrant further investigation if the practical difficulties encountered in this study can be addressed.

Another key finding of this study was the issue of the conversion of experimental flux data to K_p and the conversion of predicted K_p values to flux, to allow comparison between different models and experimental data. The measured solubility either saturated in the aqueous or the donor fluid made a marked difference in predicted permeation and the rank order. As such, this would appear to introduce another level of variance in the data, and also allowed a certain subjectivity in the presentation of results. Perhaps, in addition to those points highlighted by Cronin and Schultz[11] it should be recommended that models of percutaneous absorption are standardised to use either K_p or J_{max}. Certainly, the latter parameter is of substantially greater relevance to the in-use performance of topical pharmaceutical products. It should be noted that the model developed by Magnusson *et al*.[4] used J_{max}, yielding a relationship between it and the molecular weight of permeants.

Conclusions

The findings of this study would suggest that, when tested with a range of potential skin permeants – all of which are generally consistent with the nature and range of the datasets used to construct the models – the models failed to accurately predict permeation and in some cases were not within two- or three-orders of magnitude of the experimentally derived values. However, the models were able to effectively rank the relative permeation of the permeants which, while not entirely the point of the models, was an effective outcome for selecting an optimal lead candidate and suggested that such models could be used to refine a large number of potential drug candidates to produce a more manageable number to characterise experimentally. It could also be argued that rather than focussing solely on the physicochemical parameters of the penetrants, it may be the case that additional descriptors, including those relating to the biological matrix itself and the properties of the system in which the candidate molecule is applied are required to accurately model the absorption process. In addition, it is important to note that although such predictive models have been considered to be beneficial in providing an early estimation, prediction, and simulation of skin absorption, particularly during drug discovery, they do not take into account actual relative drug candidate potencies. The reality is that for any drug applied topically for the treatment of localised skin disease its efficacy is not only dependent upon its ability to penetrate the skin but its potency at the active site. As such the half maximal inhibitory concentration/half maximal effective concentration of the candidate molecules need to be considered. Such data demonstrated that when selecting candidate molecules for topical formulation development their ability to penetrate the skin (whether theoretical or experimentally derived), relative potencies, metabolism, solubility, stability and potential toxicity all need to be taken into account in the initial screening process.

Declarations

Conflict of interest

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

Funding

The authors would like to thank Nycomed and MedPharm Ltd for funding this work.

References

1. Scheuplein RJ, Blank IH. Permeability of the skin. *Physiol Rev* 1971; 51: 702–747.
2. Michaels AS *et al.* Drug permeation through human skin: theory and *in vitro* experimental measurement. *AIChE* 1975; 21: 985–996.
3. Potts RO, Guy RH. Predicting skin permeability. *Pharm Res* 1992; 12: 663–669.
4. Magnusson BM *et al.* Molecular size as the main determinant of solute maximum flux across the skin. *J Invest Dermatol* 2004; 122: 993–999.
5. Roberts M. Percutaneous absorption of phenolic compounds. Sydney: University of Sydney, 1976 (dissertation).
6. Abraham M *et al.* The factors that influence skin penetration of solutes. *J Pharm Pharmacol* 1995; 47: 8–16.
7. 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.
8. Moss GP *et al.* The application of Gaussian processes in the prediction of percutaneous absorption. *J Pharm Pharmacol* 2009; 61: 1147–1153.
9. Sun Y *et al.* Predictions of skin penetration using machine learning methods. In: Giannotti F, Gunopulos D, Turini F, Zaniolo C, Ramakrishnan N, Wu XD, eds. *Proceedings of 8th IEEE International Conference on Data Mining*. Pisa, Italy, 2008; 1049–1054.
10. 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.
11. Cronin MTD, Schultz TW. Pitfalls in QSAR. *J Mol Struct: Theochem* 2003; 622: 39–51.
12. Yamashita Y, Hashida M. Mechanistic and empirical modelling of skin permeation of drugs. *Adv Drug Deliv Rev* 2003; 55: 1185–1199.
13. Farahmand F, Maibach HI. Estimating skin permeability from physicochemical characteristics of drugs: a comparison between conventional models and an *in vivo* based approach. *Int J Pharm* 2009; 375: 41–47.
14. Moss GP, Cronin MTD. Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption: re-analysis of steroid data. *Int J Pharm* 2002; 238: 105–109.
15. Degim IT *et al.* Prediction of skin penetration using artificial neural network (ANN) modelling. *J Pharm Sci* 2003; 92: 656–664.
16. Neumann D *et al.* A fully computational model for predicting percutaneous drug absorption. *J Chem Inf Model* 2006; 46: 424–429.
17. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. London, England: The MIT Press, 2006.
18. Sun Y *et al.* The application of stochastic machine learning methods in the prediction of skin penetration. *Applied Soft Computing* 2011; 11: 2367–2375.
19. 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.
20. Barratt MD. Quantitative structure-activity relationships for skin permeability. *Toxicology In Vitro* 1995; 9: 27–37.
21. 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.
22. Kirchner LA *et al.* The prediction of skin permeability by using physicochemical data. *Altern Lab Anim* 1997; 25: 359–370.
23. Cronin MTD *et al.* Investigation of the mechanism of flux across human skin *in vitro* by quantitative structure-permeability relationships. *Eur J Pharm Pharmacol* 1999; 7: 325–330.
24. Wilschut A *et al.* Estimating skin permeation — the validation of 5 mathematical skin permeation models. *Chemosphere* 1995; 30: 1275–1296.
25. Prapapopolou M *et al.* Prediction of skin penetration using machine learning methods. In: Brain KR, Walters KA, eds. *Perspectives in Percutaneous Penetration. Volume 11*. Cardiff: STS Publishing, 2008: 59.
26. Fedors RF. A method for estimating both the solubility parameters and molar volumes of liquids. *Poly Eng Sci* 1974; 14: 147–154.
27. Kligman AM, Christophers E. Preparation of isolated sheets of human stratum corneum. *Arch Dermatol* 1963; 88: 702–705.
28. Chilcott R *et al.* Inter-laboratory variation of in vitro diffusion cell measurements: an international multi-centre study using quasi-standardised methods and materials. *J Pharm Sci* 2005; 94: 632–636.
29. Moss GP *et al.* The application and limitations of mathematical modelling in the prediction of permeability across mammalian skin and polydimethylsiloxane membranes. *J Pharm Pharmacol* 2011; 63: 1411–1427.
30. Moss GP *et al.* Design, synthesis and characterization of captopril produgs for enhanced percutaneous absorption. *J Pharm Pharmacol* 2006; 58: 167–177.
31. 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.
32. 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.
33. Hostynek JJ, Magee PS. Modelling in vivo human skin absorption. *Quant Struct Act Relat* 1997; 16: 473–479.
34. Lee CK *et al.* Skin permeability of various drugs with different lipophilicity. *J Pharm Sci* 1994; 4: 562–565.
35. Morimoto Y *et al.* Prediction of skin permeability of drugs: comparison of human and hairless rat skin. *J Pharm Pharmacol* 1992; 44: 634–639.
36. 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.