JPP Journal of Pharmacy and Pharmacology

JPP 2010, 62: 738–749

Received November 2, 2009
Accepted February 1, 2010
DOI 10.1211/jpp/62.06.0010
ISSN 0022-3573

Research Paper

The application of feature selection to the development of Gaussian process models for percutaneous absorption

Lun Tak Lam[a], Yi Sun[b], Neil Davey[b], Rod Adams[b], Maria Prapopoulou[a], Marc B. Brown[a] and Gary P. Moss[c]

[a]The School of Pharmacy and [b]The School of Engineering & Information Sciences, University of Hertfordshire, College Lane, Hertfordshire and [c]The School of Pharmacy, Keele University, Staffordshire, UK

Abstract

Objectives The aim was to employ Gaussian processes to assess mathematically the nature of a skin permeability dataset and to employ these methods, particularly feature selection, to determine the key physicochemical descriptors which exert the most significant influence on percutaneous absorption, and to compare such models with established existing models.

Methods Gaussian processes, including automatic relevance detection (GPRARD) methods, were employed to develop models of percutaneous absorption that identified key physicochemical descriptors of percutaneous absorption. Using MatLab software, the statistical performance of these models was compared with single linear networks (SLN) and quantitative structure–permeability relationships (QSPRs). Feature selection methods were used to examine in more detail the physicochemical parameters used in this study. A range of statistical measures to determine model quality were used.

Key findings The inherently nonlinear nature of the skin data set was confirmed. The Gaussian process regression (GPR) methods yielded predictive models that offered statistically significant improvements over SLN and QSPR models with regard to predictivity (where the rank order was: GPR > SLN > QSPR). Feature selection analysis determined that the best GPR models were those that contained log P, melting point and the number of hydrogen bond donor groups as significant descriptors. Further statistical analysis also found that great synergy existed between certain parameters. It suggested that a number of the descriptors employed were effectively interchangeable, thus questioning the use of models where discrete variables are output, usually in the form of an equation.

Conclusions The use of a nonlinear GPR method produced models with significantly improved predictivity, compared with SLN or QSPR models. Feature selection methods were able to provide important mechanistic information. However, it was also shown that significant synergy existed between certain parameters, and as such it was possible to interchange certain descriptors (i.e. molecular weight and melting point) without incurring a loss of model quality. Such synergy suggested that a model constructed from discrete terms in an equation may not be the most appropriate way of representing mechanistic understandings of skin absorption.

Keywords Gaussian process; machine learning methods; percutaneous absorption; quantitative structure–permeability relationships

Introduction

The prediction of skin absorption is of interest to many fields, including topical and transdermal drug delivery, cosmetics and risk assessment for dermal exposure. The development of viable, quantitative models has been an area of substantial interest for almost 20 years, and offers considerable advantages in reducing or replacing time-consuming and costly experiments. It is known that the physicochemical properties of a molecule exert a substantial effect on its permeability, and as such most predictive methods have relied on a qualitative or quantitative appraisal of such properties, usually as discrete entities within a mathematical representation of permeation, to understand the mechanisms of absorption and to allow prediction of the penetration of a range of exogenous chemicals. In particular, the effects of lipophilicity (most commonly expressed as log P, the octanol–water partition

Correspondence: Gary P. Moss, The School of Pharmacy, Keele University, Keele, Staffordshire ST5 5BG, UK. E-mail: gpjmoss@yahoo.co.uk

coefficient), hydrogen bonding, molecular weight (or size) and melting point were considered highly significant in their influence, and therefore in predicting permeability.[1,2] Subsequently, several researchers determined that molecular size was more significant than previously suggested.[3,4]

It is interesting to consider the nature of descriptors returned by different analyses of datasets. This is clearly highlighted by Potts and Guy.[3,5] In those two studies, the authors determined that the relationship between K_p and physicochemical descriptors differed as the nature of the dataset (from Flynn[6]) was, in the later study, qualitatively examined and abbreviated to 37 compounds. This subset was shown to be dependent on lipophilicity and hydrogen-bonding, whereas an analysis of the whole dataset demonstrated that lipophilicity and molecular weight were the key determinants in percutaneous absorption.[3]

Hydrogen bonding, despite being absent from the seminal Potts and Guy[3] model, has been considered as a key influence in percutaneous absorption for just over 30 years.[7] Partition phenomena, and in particular the development of the solvatochromic theory and developments in the understanding of epidermal permeability indicated the importance of hydrogen-bonding acceptor and donor properties in percutaneous absorption.[8–10]

Roberts *et al.*[11] showed that the introduction of even one hydrogen-bonding group to a molecule could result in a significant decrease in its permeability, whereas the addition of further groups to the molecule resulted in further, smaller, nonlinear decreases. They concluded that hydrogen bonding was the key factor in diffusion across the stratum corneum, whereas lipophilicity was more important for partitioning and may be related to the pK_a of the penetrant.

While it is difficult to directly compare the studies discussed above with other approaches (due to, for example, differences in dataset composition or statistical methods of analysis), it may be argued that the use of methods that do not properly consider the nature of the dataset used undermines any resultant model. Moss *et al.*[12] compared the statistical accuracy of Gaussian processes, single linear networks (SLN) and quantitative structure–permeability relationships (QSPRs) by a range of statistical methods, and found that the nature of the dataset was inherently nonlinear and that skin permeation (as represented by K_p) was best described, in purely statistical terms, by Gaussian process approaches.

As this field expanded, a large number of studies presented a diverse range of models based on an array of different, often complementary, datasets, and an increasing number of physicochemical properties, including hydrogen bonding and molecular size, were presented.[5,9,11,13–15] Various modifications have been made to these models, some of which involve the use of nonlinear modelling. For example, Wilschut *et al.*[16] examined five mathematical models by nonlinear multiple regression. The octanol–water partition coefficient and molecular weight were used as independent parameters. They suggested that a modified form of the Potts and Guy[3] equation best modelled skin absorption. Finally, to understand the scope, limitations and context of these models, and how they should be applied, it must be emphasised that they are all based on infinite doses being delivered from aqueous vehicles.

Therefore, while nonlinear modelling of skin absorption is not new, it is certainly an area which has not been extensively or systematically explored. The aim of this study was to compare further the statistical accuracy and predictive ability of linear and nonlinear methods of modelling, and to explore combinations of molecular descriptors that may influence, individually or synergistically, percutaneous absorption.

Materials and Methods

Dataset

The dataset employed in this study was obtained from Moss *et al.*[12] Briefly, it is a dataset that contains 142 different chemicals and their associated physicochemical descriptors and permeability values (K_p, as cm/h). It is an extension of that published by Flynn[6] and utilised in the study by Potts and Guy.[3] It is supplemented by the addition of data from previous publications and from the Edetox database (available at www.ncl.ac.uk/edetox/index.html).[15–17] It includes the data, obtained from the literature, for six physicochemical descriptors of each compound, namely, molecular weight (MW), melting point (MPt), solubility parameter (SP), the octanol–water partition coefficient (log P, used as provided in the sources listed above), hydrogen bonding acceptor groups (HA) and donor groups (HD).[18]

Mathematical methods for model development

The mathematical methods employed herein have been described in detail elsewhere.[12,19] The modelling in this study was carried out by a combination of machine learning methods and QSPRs. The QSPRs employed are those by Potts and Guy[3], Cronin *et al.*[14], Moss and Cronin[20] and Luo *et al.*[21]

Machine learning methods include SLN, which is a simple linear regression (it is the same as a linear regression method and uses iterated re-weighted least-squares training) and Gaussian process regression (GPR), which is a regression that calculates the relationship between variables via a nonlinear processes. Further, Gaussian process regression with automatic resonance detection (GPRARD) has been employed to calculate the relative significance of the molecular descriptors in GPR modelling.[12] Performance measures of GPR, SLN, GPRARD and QSPRs were calculated via Matlab R2008a. This program relies on tailored scripts (essentially, a series of commands that allow Matlab to process the required calculations) to conduct calculations for the specific tests used in this study and in previous studies i.e. Moss *et al.*[12] The scripts used were analysis by SLN/QSPR, GPR/GPRARD, GPR (improvement over the naïve model (ION), with statistical significance determined by a paired *t*-test), one script for GPR (normalised mean squared error (NMSE), paired *t*-test), GPR (correlation coefficient (CORR), r, paired *t*-test), SLN (ION paired *t*-test), SLN (NMSE paired *t*-test) and for SLN (CORR paired *t*-test). Matlab was used also to perform statistical analysis of performance measures between SLN and GPR. Statistical comparisons between QSPR and machine learning methods (GPR and SLN) were performed using SPSS (version 16).

Quantitative structure–permeability relationship analysis

Before the application of the modelling methods described below to the dataset, the QSPR methods were applied to the data to provide a comparison between machine learning methods and previous approaches to this matter. The methods used were those reported previously.[3,14,20] Further details on the nature of these models may be found elsewhere.[14,20]

Machine learning methods

Single layer networks

Regression analysis was initially carried out on the dataset using a SLN. This simple linear regression considers the output, y, as the weighted sum of the components of an input vector, x, which can be written as follows:

$$y = y(x; w) = \sum_{i=1}^{d} w_i x_i + w_0 \quad (1)$$

where d is the dimensionality of the input space (i.e. the number of features used to describe a molecule) and $w = (w_1; : : : ; w_d; w_0)$ is the weight vector. The weights are set so that the sum squared error function is minimised on a training set.

Gaussian process regression

Gaussian process modelling is a nonparametric method. It does not produce an explicit functional representation of the data, as QSPR modelling does in the form of an equation where the permeability is usually related to statistically significant physicochemical descriptors of a dataset. In GPR modelling it is assumed that the underlying function that produces the data, $f(x)$, will remain unknown, but that the data are produced from a (infinite) set of functions, with a Gaussian distribution in the function space. This has been described in detail elsewhere.[12,19] Briefly, a Gaussian process is completely characterised by its mean and covariance function. The mean function is normally considered to be the 'zero everywhere' function. The covariance function, $k(x_i, x_j)$, is crucial to Gaussian process modelling as it expresses the expected correlation between the values of $f(x)$ at the two points x_i, x_j. In other words, it defines nearness or similarity between data points. Since the model employed herein is a Gaussian process, this distribution is also Gaussian and is therefore fully defined by its mean and variance. The mean at x_* is given by:

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

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:

$$\mathrm{var}[y_*] = k(x_*, x_*) - k_*^T (K + \sigma_n^2 I)^{-1} k_* \quad (3)$$

where $k(x_*,x_*)$ denotes the variance of y_*. In this study, the mean has been used as the prediction and the variance as error bars on the prediction.

Gaussian process regression with automatic relevance determination (GPRARD)

To implement automatic relevance determination in GPR, the characteristic length-scale matrix, M, is redefined as a diagonal matrix containing the elements of vector $L = [l_1^{-2},\ldots,l_D^{-2}]$, and $l_1,\ldots,l_D$ on the diagonal are the characteristic length scales for each input dimension, determining how relevant an input is to the task.[22] If the length-scale has a very large value, it suggests that the corresponding input could be removed from the inference. These characteristic length-scales can be optimised from the data by Bayesian inference.

Feature selection

The features, or molecular descriptors, most frequently used in studies of modelling percutaneous absorption were employed in this study. Parameters were used that were readily accessible and calculable without the need for expensive, specialist software.[12,20] The features utilised in this study are listed above.

Analysis of the dataset

Data was visualised by scatter diagrams plotted with Microsoft Excel 2007, to discern patterns between the features. Such visualisation has been shown previously.[12]

The dataset was divided, for machine learning method development, into a training set and a test in the ratio of 75% (107 compounds) and 25% (35 compounds), respectively.[23] The compounds were randomly allocated into the subsets automatically by Matlab R2008a via primeSeed code, which acts as a recorder to document the allocation of the compounds in the subsets. In total, the experiment was repeated 10 times, generating 10 different test sets. Each test set contained a unique primeSeed code that recorded the compounds allocated in the corresponding test set. The same primeSeed codes were included in every script for the machine learning method and QSPR to ensure identical compounds were tested by each method.

Regression modelling was employed with each combination of descriptors as input vectors. In Gaussian process modelling, the initial values of the logarithms of the characteristic length scale, the signal variance and the noise variance were chosen using cross validation from 10 user-defined pre-sets. In addition, a fivefold cross-validation procedure was used to select optimal parameters for each test. In such cases, each training set was divided further into training and validation sets five times.

To investigate which compound descriptors contributed significantly to the prediction, GPRARD methods were applied to the dataset. Experiments were again conducted on 10 randomly selected training and test sets. However, in this case the hyperparameters were optimised by maximising the marginal likelihood using the derivative rather than selecting from pre-set hyperparameters using a cross validation procedure.[24] In each case the logarithms of characteristic length-scale, signal variance and noise variance were initialised for each input dimension, as [0; 0; 0; 0; 0; 0; log(SQRT(0.1))]. Rasmussen and Williams'[19] Gaussian Processes Toolbox was applied to the dataset to carry out Gaussian process modelling.

Performance measurements of QSPR models and machine learning methods were calculated via Matlab R2008a. The parameters employed to ascertain statistical quality of each model were percent improvement over the naïve model, (ION, %), normalised mean squared error (NMSE) and the correlation coefficient (CORR), as described above and employed previously.[12]

Results

Consideration of the results of this study can be divided broadly into four regions: the shape of the distribution between physiochemical properties and skin permeability; the comparison of the prediction accuracy between machine learning methods modelling and QSPR models; the comparison of the accuracy to quantify percutaneous absorption via nonlinear and linear approaches; and the selection of features that are significant in the mathematical quantification of skin absorption. As a measure of performance, ION, NMSE and CORR have been employed.

Distribution of the physicochemical parameters and permeability coefficients

Visualisation of the data provided an insight into the relationship between physiochemical properties and permeability coefficients among 142 compounds employed in this study. In common with previous work in this field, the visualisation of the data shown previously[12] suggested that the underlying relationship between the physicochemical descriptors and permeability coefficients was inherently nonlinear.[12,25] Moreover, other molecule descriptors including melting point, molecular weight, solubility parameters, HA and HD also showed nonlinearity with log K_p.

Further, the visualisation of the data described previously indicated that the skin permeability coefficient was not solely dependent on one molecular descriptor.[12,25] Compounds with similar properties for one particular feature can demonstrate enormous variations in log K_p. For example, if compounds with one hydrogen bond donor group are considered, log K_p is observed to vary from −1.2 to −5.0 approximately.

However, it should be noted that certain parameters, such as hydrogen bond donor and acceptor groups, may be considered as discrete rather than continuous variables, and as such a linear relationship between these parameters and descriptors that are continuous in nature (such as log P, molecular weight or solubility parameter) should not necessarily be expected, and may be of limited statistical value.

Statistical evaluation of model quality

Figure 1 indicated that those latter models, derived from a more comprehensive extension of Flynn's[6] dataset (for example, those that incorporate data from other studies) resulted in improved predictions.[26–28] It should also be noted that, as expected, the model proposed by Barratt[29] performed relatively weakly due to the limitation in the number of observations included in that study, a point made previously in the literature.[24] This suggested the importance

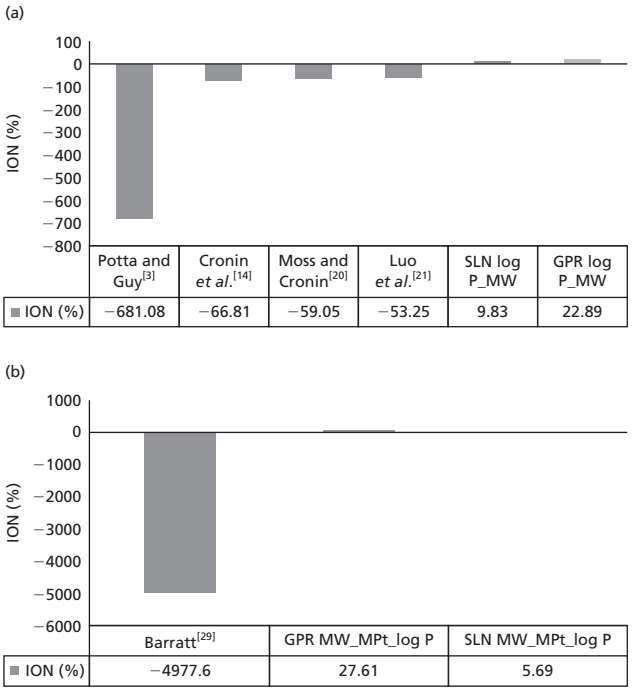

Figure 1 Relationship between physiochemical properties and permeability coefficients among 142 compounds. Comparison of the improvement over the naïve model (ION) for (a) machine learning methods (GP and SLN) compared with a range of quantitative structure–permeability relationships (QSPR) models that relate the permeability of a penetrant to the octanol–water partition coefficient (log P) and molecular weight (MW) and (b) with the QSPR model proposed by Barratt.[29] SLN, single linear networks; GPR, Gaussian process regression; MPt, melting point

of dataset validity, particularly with regard to size, the consistency of experimental protocols, reproducibility, and comprehensiveness in model developments.[20] This point was highlighted by Moss and Cronin[20] who developed a QSPR model, which did not include the steroid data used by Scheuplein *et al.*,[30] which was collated into Flynn's[6] dataset, but instead used the data colleted by Johnson *et al.*[25] The inclusion of the model by Barratt[29] suggested a possible limitation in the use of this data, which the results in Figure 1 would appear to substantiate.

As discussed above, log P, molecular weight and terms pertaining to hydrogen bonding have been widely identified as highly significant phenomena in developing a mechanistic understanding of percutaneous absorption. Despite this, the QSPR-type models employed in this study – and which contain most, if not all of these parameters – fail to accurately predict K_p; in most cases, they return predictions that are, in terms of the statistical tests used to compare the performance of models (i.e. measures of ION, NMSE and CORR), significantly worse than the naïve model, which is simply the average K_p value of the whole dataset. It can be seen that, by using the same parameters, Gaussian process and SLN models provide statistically better results than QSPR models, particularly in terms of higher ION and lower NMSE values, although difference is NMSE values are not always as pronounced as those for ION. Nevertheless, the

Table 1 Performance measures and statistical comparisons of quantitative structure–permeability relationships and corresponding machine learning methods models

Descriptors	Models or machine learning method assessed	ION (%) ± SD (higher number is best)	NMSE ± SD (smallest number is best)	CORR ± SD (highest value is best)	*P* values (QSPR vs SLN)			*P* values (QSPR vs GP)			*P* values (SLN vs GP)		
					ION (%)	NMSE	CORR	ION (%)	NMSE	CORR	ION (%)	NMSE	CORR
log P and molecular weight	Potts & Guy[3]	–681.08 ± 139.93	7.52 ± 0.80	0.25 ± 0.11	0.00	0.00	0.08	0.00	0.00	0.01			
	Cronin *et al.*[14]	–66.81 ± 41.90	1.70 ± 0.42	0.31 ± 0.12	0.00	0.00	0.36	0.00	0.00	0.00			
	Moss & Cronin[20]	–59.05 ± 39.06	1.62 ± 0.39	0.29 ± 0.12	0.00	0.00	0.26	0.00	0.00	0.002			
	Luo *et al.*[21]	–53.25 ± 36.24	1.56 ± 0.35	0.24 ± 0.11	0.00	0.00	0.06	0.00	0.00	0.00			
	SLN	9.83 ± 11.1	0.93 ± 0.17	0.34 ± 0.17							0.00	0.00	0.01
	GPR	22.89 ± 10.62	0.79 ± 0.17	0.49 ± 0.11									
log P, molecular weight and melting point	Barratt[29]	–4977.6 ± 3318.9	51.60 ± 34.32	0.30 ± 0.15	0.00	0.00	0.76	0.00	0.00	0.00			
	SLN	5.69 ± 12.32	0.97 ± 0.19	0.31 ± 0.14							0.00	0.00	0.00
	GPR	27.61 ± 9.32	0.75 ± 0.15	0.53 ± 0.11									

In all cases, 10 evaluations were carried out to measure standard deviations and *P* values for each parameter. Statistical analysis relates to *P* values for *t*-tests carried out between the performance measures of each quantitative structure–permeability relationships (QSPR) model and corresponding single layer network (SLN) machine learning models. CORR, correlation coefficient; GP, Gaussian process; GPR, Gaussian process regression; ION, improvement over the naïve model; NMSE, normalised mean squared error.

improvement is statistically valid ($P < 0.05$, Table 1). For the combination of features discussed in the preceding section, the Gaussian process demonstrated the best results, even compared with SLN, in terms of both the model's prediction accuracy (ION) and stability (NMSE), as shown in Table 1. However, as discussed previously, the nature of the dataset and its compatibility with a particular mathematical approach should be considered.[12] Nevertheless, Figure 2 shows a clear improvement in the predictivity of the Gaussian process model, compared with Potts and Guy.[3]

Feature selection

Due to the large number of statistical comparisons made between each possible combination of molecular features, Table 2 shows only a condensed comparison of the statistical tests carried out comparing the combination of features in Gaussian process models. Specifically, it includes only models that demonstrated no significant difference compared with the highest ION (%) ranked model (GPR: MPt, log P and HD), in effect, the best performing combination of features as defined by the results of the statistical comparison of models. Table 3 shows the Gaussian process models that, on both ION (%) and NMSE measurements, demonstrated no significant difference compared with Gaussian process models with better performance measures, as well as either no significant difference or significant improvement compared with Gaussian process models with worse performance measures. The statistically 'best' Gaussian models all contain the specific combination of log P and HD, coupled with either melting point or molecular weight. It appears that melting point and molecular weight are, in a purely modelling

Figure 2 Comparison of the predictive ability of Gaussian process models with the quantitative structure–permeability relationship model proposed by Potts and Guy[3] across a wide range of lipophilicities. Data points shown were obtained from a subset of the overall dataset (as described in the Materials and Methods section). The test set shown was that which resulted in the Potts and Guy[3] model achieving the best performance among the 10 test sets generated by this analysis

Table 2 Summary of the statistical analysis of the comparisons of Gaussian process models with different combinations of physicochemical descriptors

ION (%) ranking	GPR combination of features	MPt.log P.HD	MW.MPt. log P.HD	MW.MPt.SP. log P.HD	MPt.log P.HA.HD	MW.SP.log P.HD	MPt.SP.log P.HD	MW.MPt. log P.HA. HD	MW.log P.HD	MW.log P.HA.HD	MW.MPt.SP. log P.HA.HD	MW.MPt. HA	MW.MPt. HA.HD
1	GPR MPt.log P.HD	–	X	X	X	X	X	X	X	X	X	X	X
2	GPR MW.MPt. log P.2HD	X	–	X	X	X	X	X	X	√	X	X	√
3	GPR MW.MPt. SP.log P.HD	X	X	–	X	X	X	√	X	√	X	√	√
4	GPR MPt.log P.Ha.HD	X	X	X	–	X	X	X	X	X	X	X	X
5	GPR MW.SP.log P.HD	X	X	X	X	–	X	X	X	X	X	X	√
6	GPR MPt.SP.log P.HD	X	X	X	X	X	–	X	X	X	X	X	√
7	GPR MW.MPt.log P.Ha.HD	X	X	√	X	X	X	–	X	X	√	X	√
8	GPR MW.log P.HD	X	X	X	X	X	X	X	–	X	X	X	X
9	GPR MW.log P.Ha.HD	X	√	√	X	X	X	X	X	–	X	X	X
10	GPR MW.MPt. SP.log P.Ha.HD	X	X	√	X	X	X	√	X	X	–	X	√
11	GPR MW.MPt. HA	X	X	√	X	X	X	X	X	X	X	–	√
12	GPR MW.MPt. HA.HD	X	√	√	X	√	√	√	X	X	√	√	–

Note: X indicates no significant difference ($P < 0.05$) and √ indicates a significant difference ($P > 0.05$) between the two groups compared. GPR, Gaussian process regression; ION, improvement over the naïve model; MW, molecular weight; MPt, melting point; SP, solubility parameter; log P, the octanol–water partition coefficient; HA, hydrogen bonding acceptor groups; HD, hydrogen bonding donor groups.

Table 3 Statistical performance measures of the best-performing models, and significance of molecular descriptors employed in the Gaussian process models

Combination of features	Statistical performance measures			Length scale						
	ION (%) ± SD	INMSE ± SD	CORR ± SD	MW	MPt	SP	log P	HA	HD	Features significance ranking
MPt.log P.HD	37.59 ± 8.54	0.64 ± 0.13	0.63 ± 0.09	–	1.23	–	0.51	–	0.99	log P > HD > MPt
MW.MPt.log P.HD	37.40 ± 7.56	0.65 ± 0.15	0.62 ± 0.09	5.22	1.28	–	0.51	–	1.03	log P > HD > MPt > MW
MW.MPt.SP.log P.HD	37.35 ± 7.23	0.65 ± 0.14	0.62 ± 0.09	5.20	1.27	31.09	0.51	–	1.0	log P > HD > MPt > MW > SP
MPt.log P.Ha.HD	35.19 ± 10.81	0.67 ± 0.18	0.62 ± 0.10	–	1.14	–	0.85	2.51	1.11	log P > HD > MPt > HA
MW.SP.log P.HD	35.12 ± 7.08	0.67 ± 0.12	0.62 ± 0.08	0.77	–	83.70	0.64	–	0.62	HD > log P > MW > SP
MPt.SP.log P.HD	34.21 ± 11.46	0.68 ± 0.19	0.61 ± 0.10	–	1.22	24.47	0.51	–	0.98	log P > HD > MPt > SP
MW.log P.HD				0.77	–	–	0.64	–	0.62	HD > log P > MW
MW.log P.Ha.HD				0.62	–	–	0.78	0.64	0.41	HD > HA > log P > MW
MW.MPt.SP.log P.Ha.HD				0.90	1.31	53.92	0.86	0.70	0.39	HD > HA > log P > MW > MPt > SP
MW.MPt.Ha				0.38	0.86	–	–	0.43	–	MW > HA > MPt
MW.MPt.log P.Ha.HD				0.90	1.32	–	0.87	0.70	0.40	HD > HA > log P > MW > MPt
MW.MPt.Ha.HD				0.26	1.91	–	–	0.38	0.70	MW > HA > HD > MPt

CORR, correlation coefficient; ION, improvement over the naïve model; MW, molecular weight; MPt, melting point; NMSE, normalised mean squared error; SP, solubility parameter; log P, the octanol–water partition coefficient; HA, hydrogen bonding acceptor groups; HD, hydrogen bonding donor groups.

context, interchangeable in this process, and replacing one with the other does not exert a detrimental effect on a particular model. It should also be pointed out that the reduced correlation coefficient observed for the Potts and Guy[3] equation (in Table 1) may have been as a result of the application of this model to our dataset, which differed from that used originally to develop this model, and which may be potentially of limited value. Table 4 shows a summary of length scale analysis, calculated by Gaussian process automatic resonance detection for each feature in the Gaussian process models recorded in Table 2. Essentially, a lower length scale value indicated a higher significance of the role of a particular molecular descriptor in predictions of permeability coefficients. From Table 4, it could be seen that the difference in the length scale factor between the molecular features in each model was relatively small. The only exception was the solubility parameter, which demonstrated a minimum of two significance figures difference compared with other molecular descriptors. In essence, this indicated that the solubility parameter was not a significant feature in the quantification of percutaneous absorption. This was further supported by addition of solubility parameters into the combination of descriptors, which did not lead to a significant improvement in model predictivity. For example, the Gaussian process combinations (GPR: MW, MPt, SP, log P, HD) and (MPt, log P, HA, HD) offered equally significant predictions of log K_p. In some cases the inclusion of solubility parameters could cause significant reductions in the predictivity of a model, for example, the Gaussian process combination GPR: MW, MPt, log P, HA, HD was more significant than GPR: MW, MPt, SP, log P, HA, HD.

The results of this particular analysis suggested that, using the physicochemical descriptors of log P, the number of hydrogen bonding donor groups, and either molecular weight or melting point, resulted in a Gaussian process model with optimal predictivity and that the addition of further molecular descriptors did not improve the quality of the model.

Comparison of nonlinear and linear predictions of skin absorption

Lian *et al.*[31] suggested that the simplicity of linear equations enhanced the ability of a model to provide accurate predictions. This comment has been explored in this study, where the difference in predictivity of the permeability coefficient between Gaussian process and SLN modelling has been examined. The results in Table 3 indicated that the Gaussian process provided significantly better predictions of

Table 4 Statistical analysis of the best models obtained by Gaussian process, single layer network and quantitative structure–permeability relationships methods

Model 1	Model 2	Model 1		Model 2		*P* value		Significant difference
		ION (%) ± SD	NMSE ± SD	ION (%) ± SD	NMSE ± SD	ION (%)	NMSE	
GPR MPt.log P.HD	SLN MPt.HA	37.59 ± 8.54	0.64 ± 0.13	11.23 ± 11.29	0.91 ± 0.13	0.00	0.00	Y
GPR MPt.log P.HD	Luo *et al.*[21]	37.59 ± 8.54	0.64 ± 0.13	–53.25 ± 36.24	1.56 ± 0.35	0.00	0.00	Y
SLN MPt.HA	Luo *et al.*[21]	11.23 ± 11.29	0.91 ± 0.13	–53.25 ± 36.24	1.56 ± 0.35	0.00	0.00	Y

GPR, Gaussian process regression; ION, improvement over the naïve model; MPt, melting point; log P, the octanol–water partition coefficient; HA, hydrogen bonding acceptor groups; HD, hydrogen bonding donor groups; NMSE, normalised mean squared error; SLN, single linear networks.

log K_p than SLN for the overall highest ION model, as well as the best models within its categories based on specific combinations of physicochemical descriptors. The only exception was the model with two features, where the overall best SLN model (where MPt and HA were returned as the most significant parameters) demonstrated no significant difference with the Gaussian process model (GPR: MW and HD). The results of the statistical comparisons (paired *t*-tests) of these models has been summarised in Table 4. These results suggested that Gaussian process modelling was, in statistical terms, the most appropriate model of those analysed to employ in predicting percutaneous absorption, with the observed differences being statistically significant. In terms of model quality i.e. accuracy of prediting K_p, the statistical comparisons used in this study would suggest the following rank order: Gaussian process > SLN > QSPR (all types).

Discussion

An important point in this study was that the composition of the dataset (the inputs) clearly affected the nature of any model derived (the output). This may seem obvious but it is important to make such a point, given that the dataset used in this study was different from those employed to develop the established QSPR models. However, the specific composition of the dataset can clearly influence the nature of the model. Moss *et al.*[12] discussed this, in terms of the breadth of the Flynn[6] dataset, which underpins so much of the work in this field. That dataset was composed predominantly of molecules which, for example, had log P values less than 2.0. Moss *et al.*[12] argued that this may in effect be providing only a limited picture of percutaneous absorption, limiting the applicability of the model, and this indeed has been addressed by other researchers, where nonlinear modifications of the Potts and Guy[3] equation were proposed.[16] This suggested that a simple linear relationship between K_p and any number of molecular descriptors may not fully represent percutaneous absorption, and may result in limited or inaccurate predictivity for a particular model. In addition, data visualisation also suggested a clear nonlinear relationship between physicochemical properties of molecules, suggesting no clear linear trend between any of these descriptors.[32] Those QSPR models that could be loosely described as being of the 'Potts and Guy' type suggested that a linear response existed between, for example, K_p and log P. As discussed recently, it should be noted that such a relationship only exists within the specific range of the models; ostensibly, this reflects the range of data employed to construct the model.[17,33,34] It may be the case that the models are therefore limited by the range of their dataset and that this study, and those like it, yield models that are more representative of percutaneous absorption across a wider range of physicochemical properties.

A range of nonlinear methods have been employed to improve predictions of skin absorption. Artificial neural networks (ANN) have been employed, showing high predictive power.[35] However, it is a limited method in that ANNs 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.[33] This results in idiosyncratic results, particularly as the output will tend to fit the noise in such cases, providing poor predictivity for new compounds.[36] Gaussian process methods do not alleviate all these issues, but minimise them, providing better predictions of percutaneous absorption than existing models.[12,19]

Therefore, this study employed Gaussian process methods of analysis and, in particular, Gaussian process automatic resonance detection. This measures the covariance and length scale of each feature in the combination. The inverse of the length scale determines the relevance between input and output, thereby a low length scale value implies that the input and covariance are highly dependent on each other. In other words, this can reduce the limitation of Gaussian process caused by a 'black box' approach and provide an insight into the significance of specific molecular descriptors.[12] SLNs were also evaluated as they allow interpretation of the predictivity limitations in linear model at different ranges of features, providing a comparison between linear QSPR and machine learning methods.[37]

Data visualisation also indicated that K_p was not solely dependent on one molecular descriptor.[12] Compounds with similar properties for one particular descriptor could demonstrate enormous variations in log K_p. For example, for compounds with one hydrogen bond donor group, log K_p was observed to vary from −1.2 to −5.0 approximately. Such a visualisation of data clearly demonstrated the synergic effects between the physicochemical features investigated in this study and would indicate either that more than one physicochemical descriptor was required to successfully model percutaneous absorption, or that such parameters were not independent of each other (such as the relationship between log P and molecular weight) and that the use of particular parameters may be limited in terms of gaining specific understandings of mechanisms of absorption. Further, effects such as ionisation (and therefore solubility and speciation) have not been considered by any of these studies.

Nevertheless, Figure 2 demonstrates a clear improvement in predictivity by the Gaussian process model compared with the Potts and Guy[3] model. Figure 2 contains data points obtained from a subset of the overall dataset, due to the methods employed for the generation of tests sets, as described in the previous section. The test set shown in Figure 2 was that which resulted in the Potts and Guy[3] model achieving the best performance among the 10 test sets generated by this analysis. This was compared with experimental log K_p and predicted K_p from the model with the highest ION (%) value (GPR: MPt, log P and HD). It is a good example of how Gaussian process methods provide a better fit to experimental log K_p in contrast to Potts and Guy.[3] Even with such a subset, where performance is, in effect, at an optimum, the statistical performance of the Potts and Guy[3] model results in a majority of the predicted log K_p values being distinctly different from the experimental log K_p values, as indicated by comparatively poor ION and NMSE values. Even in this case the Gaussian process model was, in statistical terms, more accurate. Further, and rather qualitatively, it may be suggested that the scatter of the output shown in Figure 2 from the Gaussian process was substantially less linear than

that from the QSPR model, and that the latter appeared to be more representative of the scatter associated with the experimental data.

In the best Gaussian process models (shown in Table 5), every combination of features contained log P and HD together with either melting point or molecular weight. This suggested that molecular weight, melting point, log P and HD were important features in permeability coefficient predictions. It also suggested at the inter-relationship and lack of independence of certain descriptors. Interestingly, in this type of combination, melting point and molecular weight were inter-exchangeable to give predictions with no significant differences; for example, the Gaussian process combinations [MPt, log P, HD] and [MW, log P, HD] produce models of a similar statistical quality. Furthermore, addition of molecular weight or melting point to these models did not significantly influence log K_p predictions, such as the Gaussian process combinations [MW, log P, HD], [MPt, log P, HD] and [MW, MPt, log P, HD], which all demonstrated no significant difference in performance measures. This inter-exchangeability implied that a high level of correlation existed between melting point and molecular weight.[38] On the other hand, it should be considered that, while molecular weight and melting point were interchangeable for modelling purposes, this did not necessarily indicate a degree of correlation between the two parameters. This was reflected in the findings of a previous Gaussian process study.[12]

HD only appeared to exert its importance in skin permeability when coupled with log P. When log P was absent, inclusion of HD in the model could significantly decrease a model's predictive power. However, when a model was constructed containing log P, HD constantly demonstrated a lower length scale value than HA. In this case, addition of HA to the model did not result in improvements in predictivity. This meant that the Gaussian process models [GPR: MW, log P, HA, HD] and [GPR: MW, logP, HA] had a similar statistical performance, whereas removal of HD could significantly reduce performance measures. For example, [GPR: MPt, log P, HD, HA] was significantly better than [GPR: MPt, log P, HA] ($P = 0.0022$).

It was not just the removal of HD that impacted on the statistical quality of models. For example, in the absence of log P, the Gaussian process model [GPR: MW, MPt and HA] demonstrated no significant difference in ION (%) value compared with [GPR: MPt, log P, HD], the latter being the model with the best overall performance measures. This might have been due to the molecular weight bias of the dataset employed in this study, which was predominantly based on Flynn.[6,39] This highlighted that the effects of ionisation might not have been considered in the development of these models, or in previous models that employed such literature data. It should be noted, however, that [GPR: MW, MPt and HA] performed poorly in NMSE measurements, indicating that this model was not, in a statistical sense, a stable and reliable combination.

The results presented in Table 5 indicated that the Gaussian process provided significantly better predictions of log K_p than SLNs for the overall highest ION model, as well as the best models within its categories based on specific combinations of physicochemical descriptors. The only exception was the model with two descriptors, where the overall best SLN model (which returned MPt and HA as being the most significant parameters) demonstrated no significant difference with the Gaussian process model ([MW and HD]).

The 'black box' approach, as presented previously, did not allow the elucidation of mechanistic information, only predictions of K_p for chemicals of interest.[12] This study, and the use of feature selection methods, allowed all combinations of molecular descriptors to be assessed for their ability to improve statistically the quality of models generated.

Table 5 Statistical analysis (paired *t*-test) between the Gaussian process and single layer network models with the highest ION (%) in each number of molecular descriptor categories

Highest ION (%) model	GPR models	SLN models	GPR ION (%) ± SD	GPR NMSE ± SD	SLN ION (%) ± SD	SLN NMSE ± SD	*P* value (ION %)	*P* value (NMSE)	Significant difference (ION %)	Significant difference (NMSE)
Overall	MPt.log P.HD	MPt.HA	37.59 ± 8.54	0.64 ± 0.13	11.23 ± 11.29	0.91 ± 0.13	0.00	0.00	Y	Y
2 features	MW.HD	MPt.HA	25.54 ± 12.90	0.77 ± 0.19	11.23 ± 11.29	0.91 ± 0.13	0.34	0.046	N	Y
3 features	MPt.log P.HD	MPt.SP. HA	37.59 ± 8.54	0.64 ± 0.13	10.77 ± 11.52	0.91 ± 0.14	0.00	0.00	Y	Y
4 features	MW.MPt. log P.HD	MW.MPt. SP.HA	37.40 ± 7.56	0.65 ± 0.15	9.36 ± 11.20	0.93 ± 0.18	0.00	0.00	Y	Y
5 features	MW.MPt. SP.log P.HD	MW.MPt. SP.HA.HD	37.35 ± 7.23	0.65 ± 0.14	6.90 ± 13.33	0.96 ± 0.19	0.00	0.00	Y	Y
6 features	MW.MPt. SP.log P.HA.HD	MW.MPt. SP.log P. HA.HD	31.61 ± 10.70	0.71 ± 0.15	3.47 ± 14.24	0.99 ± 0.20	0.00	0.00	Y	Y

CORR, correlation coefficient; GPR, Gaussian process regression; ION, improvement over the naïve model; NMSE, normalised mean squared error; SLN, single linear networks. MW, molecular weight; MPt, melting point; SP, solubility parameter; log P, the octanol–water partition coefficient; HA, hydrogen bonding acceptor groups; HD, hydrogen bonding donor groups.

While this has resulted in a clear understanding of the models that will improve prediction of K_p, it has demonstrated that the combination of descriptors responsible for such improvements is not always clear or consistent. This essentially demonstrated the interconnection of the parameters used. For example, an increase in lipophilicity can be achieved by increasing molecular weight and such increases were not necessarily linear.[40] The lipophilicity of a compound is determined by its chemical structures and the position of the aromatic ring; carbons, benzene rings and amide groups can increase log P.[41,42] As molecular weight increases, the number of carbon skeletons increases and therefore the lipophilic surface of a compound increases. The increase in number of hydrophobic alkane groups is considered as the major contribution to the increase in lipophilicity and hence, to an extent, permeability.[10] Water prefers to interact with hydrogen bonding groups or ionic molecules rather than nonpolar compounds.[38] Ghasemi and Saaidpour[43] highlighted that, as molecular weight increased, the increase in lipophilicity resulted in the compound becoming nonpolar, increasing solubility in the stratum corneum and reducing solubility in the aqueous environment (dermis). This was consistent with the findings of previous studies in this field.[12,17]

As the number of hydrogen bonding groups on a molecule increases, the ability of the molecule to form hydrogen bonds with water increases and therefore lipophilicity decreases. Therefore, a hydrogen bond can, indirectly, be an indication of log P. Fitzpatrick *et al.*[44] suggested that a hydrogen bond-related descriptor should be included in a model when there is the absence of a parameter directly relating to lipophilicity.

Compared with log P and molecular weight, the exact mechanistic understanding of how hydrogen bonding influences percutaneous absorption is less clear. Several authors have demonstrated that hydrogen bonding was highly related to skin permeability.[9,13,45] Poulin and Krishnan[46] and Potts and Guy[5] suggested that hydrogen bonding could significantly influence percutaneous absorption by reducing the ability of a compound to penetrate the skin, and that hydrogen bond acceptor groups play a more significant role than donor groups, a suggestion supported by Pugh *et al.*[13]

However, the findings of this study suggested a different conclusion, where generally acidic hydrogen bond donor groups have been shown to be more significant than generally basic hydrogen bond acceptor groups. These findings were in agreement with those presented by El Tayar *et al.*[47] and Geinoz *et al.*[41] This discrepancy may have been due to the role ionisation played in both the overall process of percutaneous absorption and in the nature of the descriptors. Poulin and Krishnan[46] suggested that the effects of lowering K_p by hydrogen bonding were particularly strong when the molecule had two or more hydrogen bonding donor or acceptor groups in the compounds. According to Roberts *et al.*[10] and Ghafourian and Fooladi,[32] inclusion of one hydrogen bond group (either a donor or acceptor) to the hydrocarbon skeleton would cause a substantial reduction in K_p. Addition of subsequent groups also reduced K_p, but did so in a nonlinear additive manner.

Hadgraft[48] highlighted that interactions between compound and the polar head groups of skin lipids in the intercellular channels play a significant role in percutaneous absorption. Molecules containing hydrogen bond groups can associate with the immobilised polar head groups of the lipids. As a result, their passage across the skin may be hindered, decreasing the diffusion coefficient and reducing their ability to diffuse across stratum corneum.[49] This may result in hydrogen bonding and ionic forces modification, which implies a change of head group domains. This complicates skin penetration as such an alteration may influence permeation of other exogenous chemicals in, for example, the same formulation. The number of hydrogen bond groups may vary during the partitioning process. For example, once a donor group donates a hydrogen bond, it has the potential to become a hydrogen bond acceptor, while groups that have not been ionised remained as hydrogen bond donor groups. Further, intermolecular hydrogen bonding has a substantial influence on aqueous solubility since the O-H and N-H bonds are strongly polarised and may readily facilitate donation.[50]

Most of the permeants in the data set are either weak acids or weak bases. Hence, ionisation can occur at different pH values.[48] According to Aberg *et al.*,[51] the skin surface is acidic with pH ranging from 4 to 6. However, the pH of extracellular fluid in the body is approximately 7.4, and implies a large pH gradient between the stratum corneum and underlying tissues. Removal or addition of a hydrogen bond can lead to a compound becoming ionised. The extracellular stratum corneum lipid contains free fatty acids that can undergo dissociation, resulting in a negative surface charge caused by the presence of ionised carboxyl groups.[51] As the skin is a negatively-charged membrane, this electrostatic interaction becomes a hindrance of ionised penetrants.[52] Thus, ionic compounds, particularly cations, have a lower ability to penetrate the skin compared with neutral compounds.

The disparity between the findings of Potts and Guy[5] and this study also relates to ionisation. During the process of experimentally measuring K_p, the solute was placed in a solvent where it was possible for the solute to interact with the solvent and, depending on the pK_a of the solute and the pH of the solvent, for the solute to ionise. It should be considered that log P values were also measured with ionisable compounds under conditions that may have favoured more ionic species, thus influencing the log P value obtained. Thus, experimental measurements of K_p might not appropriately reflect the effects of hydrogen bonding but may instead reflect the effects of ionisation.

Conclusions

In comparing different approaches for developing predictive models of percutaneous absorption, this study agreed with previous work suggesting the inherently nonlinear nature of the skin data set used.[12] Further, Gaussian process machine learning methods produced statistically more robust models than other approaches (SLN or QSPR-based models). The use of feature selection enabled the development of a mechanistic understanding of percutaneous absorption. While this approach resulted in specific models that were statistically superior, it also indicated clearly the

interdependence of the physicochemical descriptors employed in this, and in many other, studies. This suggested that the approach of quantifying models of skin absorption by means of a simple equation may have limited mechanistic value. While hydrogen bonding appeared to play an important role in percutaneous absorption, the issue of ionisation may have limited the validity and accuracy of models.

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. 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. 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.
6. 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.
7. Roberts M. Percutaneous absorption of phenolic compounds. Sydney, Australia: University of Sydney, 1976 (dissertation).
8. 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.
9. Abraham M *et al*. The factors that influence skin penetration of solutes. *J Pharm Pharmacol* 1995; 47: 8–16.
10. 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 Pharmaceutics* 1995; 126: 219–233.
11. 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 Pharmaceutics* 1996; 132: 23–32.
12. Moss GP *et al*. The application of Gaussian processes in the prediction of percutaneous absorption. *J Pharm Pharmacol* 2009; 61: 1147–1153.
13. 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 Pharmaceutics* 1996; 138: 149–165.
14. 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.
15. Patel H *et al*. Quantitative structure-activity relationships (QSARs) for the prediction of skin permeation of exogenous chemicals. *Chemosphere* 2002; 48: 603–613.
16. Wilschut A *et al*. Estimating skin permeation – the validation of 5 mathematical skin permeation models. *Chemosphere* 1995; 30: 1275–1296.
17. Moss GP *et al*. Design, synthesis and characterization of captopril produgs for enhanced percutaneous absorption. *J Pharm Pharmacol* 2006; 58: 167–177.
18. Fedors RF. A method for estimating both the solubility parameters and molar volumes of liquids. *Poly Eng. Sci* 1974; 14: 147–154.
19. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press, 2006.
20. Moss GP, Cronin MTD. Quantitative structure-permeatibility relationships (QSPRs) for percutaneous absorption: re-analysis of steroid data. *Int J Pharm* 2002; 238: 105–109.
21. Luo *et al*. Predicting human skin absorption of chemicals: development of a novel quantitative structure activity relationship. *Toxicol Ind Health* 2007; 23: 39–45.
22. Neal RM. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics No. 118. New York: Springer-Verlag, 1996.
23. Katritzky AR *et al*. Skin permeation rate as a function of chemical structure. *J Med Chem* 2006; 49: 3305–3314.
24. Geinoz S *et al*. Quantitative structure-permeation relationships (QSPeRs) to predict skin permeation: a critical evaluation. *Pharm Res* 2004; 21: 83–92.
25. Sun Y *et al.* Predictions of skin penetration using machine learning methods. In: Giannotti F *et al*., eds. *Proceedings of the 8th IEEE International Conference on Data Mining*. Pisa, Italy. ISBN 978-0-7695-3502-9, 2008.
26. Johnson ME *et al*. Permeation of steroids through human skin. *J Pharm Sci* 1995; 84: 1144–1146.
27. Kirchner LA *et al*. The prediction of skin permeability by using physicochemical data. *Alternatives to Laboratory Animals* 1997; 25: 359–370.
28. Degim IT *et al*. Skin permeability data: anomalous results. *Int J Pharm* 1998; 170: 129–133.
29. Barratt MD. Quantitative structure-activity relationships for skin permeability. *Toxicol In Vitro* 1995; 9: 27–37.
30. Scheuplein RJ *et al*. Percutaneous absorption of steroids. *J Invest Dermatol* 1969; 52: 63–70.
31. Lian G *et al*. An evaluation of mathematical models for predicting skin permeability. *J Pharm Sci* 2008; 97: 584–598.
32. Ghafourian T, Fooladi S. The effect of structural QSAR parameters on skin penetration. *Int J Pharm* 2001; 217: 1–11.
33. Neumann D *et al*. A fully computational model for predicting percutaneous drug absorption. *J Chem Inf Model* 2006; 46: 424–429.
34. Pannier AK *et al.* Fuzzy modelling of skin permeability coefficients. *Pharm Res* 2003; 20: 143–148.
35. Degim IT *et al.* Prediction of skin penetration using artificial neural network (ANN) modelling. *J Pharm Sci* 2003; 92: 656–664.
36. Guha R, Jurs PC. Interpreting computational neural network QSAR models: a measure of descriptor importance. *J Chem Inf Model* 2005; 45: 800–806.
37. Gramatica P *et al*. Statistical external validation and consensus modeling: a QSPR case study for K_{oc} prediction. *J Mol Graph Model* 2006; 25: 755–766.
38. Williams AC. *Transdermal and Topical Drug Delivery.* London: Pharmaceutical Press, 2003.
39. Poda GI *et al*. Random sampling or 'random' model in skin flux measurements? [Commentary on: Investigation of the

mechanism of flux across human skin in vitro by quantitative structure-permeability relationships]. *Eur J Pharm Sci* 2001; 14: 197–200.
40. Buchwald P, Bodor N. A simple, predictive, structure-based skin permeability model. *J Pharm Pharmacol* 2001; 53: 1087–1098.
41. Geinoz S *et al*. Quantitative structure-permeation relationships for solute transport across silicone membranes. *Pharm Res* 2002; 19: 1622–1629.
42. Refsgaard HHF *et al*. *In silico* prediction of membrane permeability from calculated molecular parameters. *J Med Chem* 48, 805– 811.
43. Ghasemi J, Saaidpour S. Quantitative structure–property relationship study of *n*-octanol–water partition coefficients of some diverse drugs using multiple linear regression. *Anal Chim Acta* 2007; 604: 99–106.
44. Fitzpatrick D *et al*. Modelling skin permeability in risk assessment – the future. *Chemosphere* 2004; 55: 1309–1314.
45. Ravesky OA, Schaper KJ. Quantitative estimation of hydrogen bond contribution to permeability and absorption processes of some chemicals and drugs. *Eur J Med Chem* 1998; .33: 799–807.
46. Poulin P, Krishnan K. Molecular structure-based prediction of human abdominal skin permeability coefficients for several organic compounds. *J Toxicol Environ Health* 2001; 62: 143–159.
47. El Tayar N *et al*. Percutaneous penetration of drugs: a quantitative structure-permeability relationship study. *J Pharm Sci* 1991; 80: 744–749.
48. Hadgraft J. Skin deep. *Eur J Pharm Biopharm* 2004; 58: 291–299.
49. Pugh WJ *et al*. Epidermal permeability–penetrant structure relationships: 4, QSAR of permeant diffusion across human stratum corneum in terms of molecular weight, H-bonding and electronic charge. *Int J Pharmaceutics* 2000; 197: 203–211.
50. Yin C *et al*. Prediction and application in QSPR of aqueous solubility of sulfur-containing aromatic esters using GA-based MLR with quantum descriptors. *Water Research* 2002; 36: 2975–2982.
51. Aberg C *et al*. Transport processes in responding lipid membranes: a possible mechanism for the pH gradient in the stratum corneum. *Langmuir* 2008; 24: 8061–8070.
52. Raiman J *et al*. Drug adsorption in human skin: a streaming potential study. *J Pharm Sci* 2003; 92: 2366–2372.