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

The influence of diffusion cell type and experimental temperature on machine learning models of skin permeability

Parivash Ashrafi[a], Yi Sun[a], Neil Davey[a], Simon C. Wilkinson[b] and Gary P. Moss[c]

[a]The School of Computing, University of Hertfordshire, Hatfield [b]Wolfson Unit, Medical School, Medical Toxicology Centre, University of Newcastle-upon-Tyne, Newcastle-upon-Tyne and [c]The School of Pharmacy, Keele University, Keele, UK

Keywords
dataset design; flow-through diffusion cells; Franz diffusion cells; machine learning; percutaneous absorption

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

Received August 8, 2019
Accepted October 26, 2019

doi: 10.1111/jphp.13203

Abstract

Objectives The aim of this study was to use Gaussian process regression (GPR) methods to quantify the effect of experimental temperature (T_{exp}) and choice of diffusion cell on model quality and performance.

Methods Data were collated from the literature. Static and flow-through diffusion cell data were separated, and a series of GPR experiments was conducted. The effect of T_{exp} was assessed by comparing a range of datasets where T_{exp} either remained constant or was varied from 22 to 45 °C.

Key findings Using data from flow-through diffusion cells results in poor model performance. Data from static diffusion cells resulted in significantly greater performance. Inclusion of data from flow-through cell experiments reduces overall model quality. Consideration of T_{exp} improves model quality when the dataset used exhibits a wide range of experimental temperatures.

Conclusions This study highlights the problem of collating literature data into datasets from which models are constructed without consideration of the nature of those data. In order to optimise model quality data from only static, Franz-type, experiments should be used to construct the model and T_{exp} should either be incorporated as a descriptor in the model if data are collated from a range of studies conducted at different temperatures.

Introduction

Quantitative models of skin permeation are normally constructed by the collation of data from disparate literature sources, irrespective of differences in the methods used to generate data. This has historically been used as the main method of dataset construction due to the paucity of available data. For example, the dataset collated by Flynn[1] contains 97 permeability values, mostly from *in vitro* skin permeation studies, for 94 different chemicals. These data were collected from 15 different literature sources,[2–17] inferring that the data might inevitably exhibit a high degree of experimental variation due to interlaboratory variation, including the use of skin from different sources and anatomical sites. This dataset is still the basis for the majority of work in this field, although it is important to note that a key message from Flynn's[1] study was that any such model is, by its very nature, an approximation of 'real-world' skin permeation and that it should always be treated as a work in progress.

Following Flynn's work, a wide range of models estimating skin permeation were published. Many used Flynn's dataset, or subsets abstracted from it, and others added to the original dataset. These studies are reviewed in detail elsewhere,[18,19] and they generally suggested that the inherent biological variation present in skin permeation data is so significant that differences in experimental methodologies from which models are built are impossible to decouple from each other.

In recent years, the volume of data from which models of skin permeation can be built has increased significantly. However, in the construction of a valid and relevant model, it is never a straightforward solution to simply use more and more data as it becomes available. Data have often been added to datasets from studies which were conducted at different temperatures, carried out using different types of

diffusion cells (either static, 'Franz-type' or flow-through, 'Bronaugh-type', illustrated in Figure 1), using different receptor compartment fluids (including a range of solvents or solutions buffered at a range of different pH values), or under occluded or unoccluded conditions. However, when models were constructed, few, if any, of these parameters were considered[18,19] usually as the variance associated with the outcome of the modelling process was attributed to the inherently variable nature of the skin and its permeability properties.

The availability of a range of software packages that allow the generation of thousands of molecular descriptors has also been used without generally considering the impact of practices on model quality. This issue was explored by building models with in excess of 2000 molecular descriptors of each member of their dataset.[20] By using a feature selection method, it was determined that the best model used only 27 of these descriptors, with many being removed as they were redundant due to being highly correlated or only having one possible value. This study recommended that a relevant dataset must contain a diverse and balanced set of chemicals and that the underlying nature of the dataset must be considered when conducting such studies.

The application of machine learning methods to this field has allowed more complex methods to be used to build models. For example, methods based on Gaussian

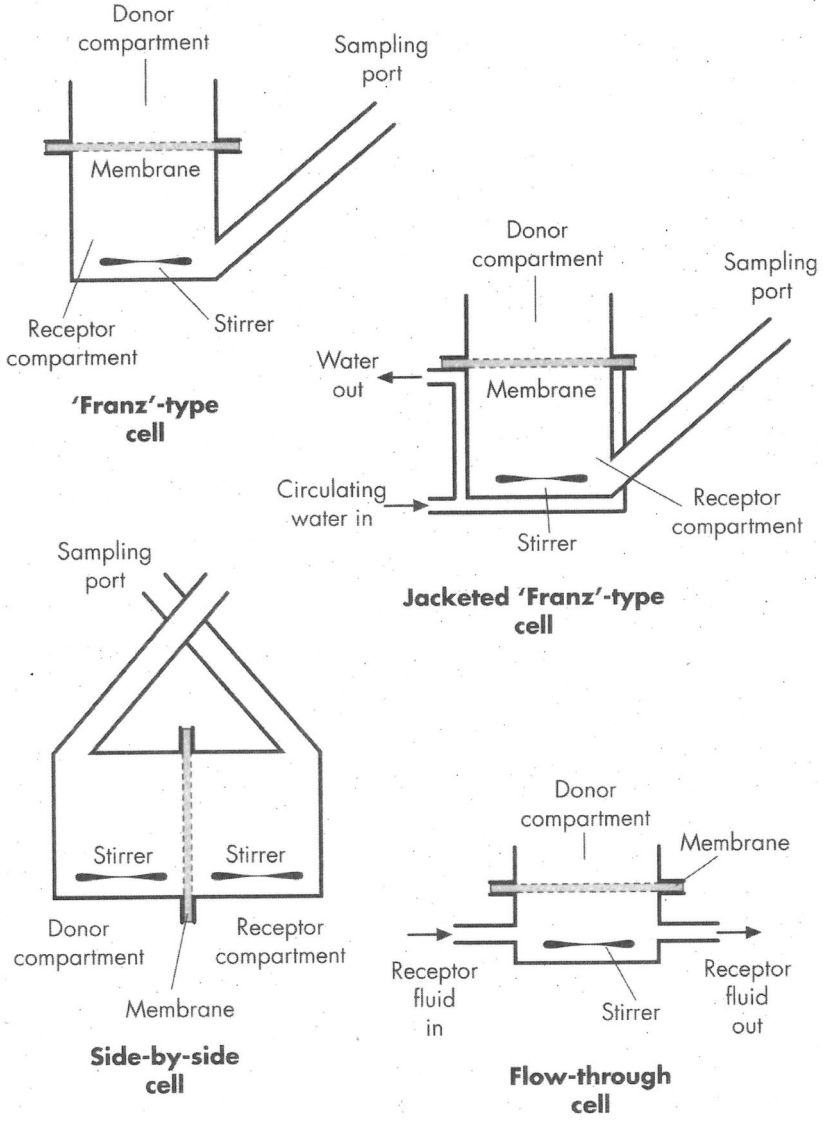

Figure 1 Examples of common diffusion cell designs, with the 'Franz'-type static diffusion cell shown top left and the flow-through cell design shown bottom right. Modified from [40], used with permission.

process regression (GPR) have demonstrated improved predictive ability compared to the more traditional quantitative structure–permeability relationship (QSPR) models.[18,21,22] These methods are not without criticism, but the often-cited absence of an algorithm to describe the permeability process is increasingly being perceived not as a limitation, but as a static outcome for such studies which may be open to incorrect interpretation, notably in the context of Flynn's comment, above.[19]

We believe that this is the first study to focus on the experimental design in the development of predictive models of skin permeation, rather than solely focusing on the physicochemical descriptors of the molecules in the dataset. This is also the first study to apply the t-SNE methodology to the modelling of skin permeability. It is therefore the aim of this study to use GPR methods to quantify the effect of various experimental conditions – specifically, experimental temperature and choice of diffusion cell (as static or flow-through) – on the development of predictive models of skin permeation. Thus, in modelling the experimental temperature and cell type, and examining the effect of these conditions on the outcome of models, this study is the first of its kind and it presents a new approach to modelling skin permeability that will have significant implications for how such models should be constructed and interpreted. Implicit in this is the characterisation of the underlying data from static and flow-through cell experiments and how this novel approach can be used to underpin better model development through rational dataset construction.

Methods

Part one – the influence of diffusion cell type on models of skin permeability

Datasets

The dataset used for the main study of diffusion cell type has been published in full[23] and variously analysed previously.[20,23] It consists of data collated from a range of literature sources based mainly on the Flynn dataset and subsequent modifications.[24–27] The dataset used in this study consists of 91 compounds from static diffusion cell studies and 53 compounds from flow-through cell experiments which are the averages of available literature data for each chemical.[21,28] The physicochemical descriptors used in the model are lipophilicity (as log P), molecular weight (MW), molar refractivity (MR), used as a measure of molecular polarizability, counts of hydrogen bond donor and acceptor groups (HD and HA, respectively), the solubility parameter (SP), described by Fedors[29]) and the melting point (MPt). The corresponding target, skin permeability coefficient, is denoted as log K_p.

Characterisation of the dataset

Data visualisation matrices and principal component analysis (PCA) have been used in previous studies on percutaneous absorption.[21,28,30] The t-distributed stochastic neighbour embedding (t-SNE) technique has recently been used for dimensionality reduction and the visualisation of high-dimensional datasets.[31] It is applied here for the first time to a dataset of skin permeability data.

Gaussian process regression

The GPR methods used have been described previously in substantial detail,[21,30] including a guide for nonexpert users to apply GPR methods to their datasets.[32] GPR is a nonparametric method of analysis. In contrast to quantitative structure-activity methods (QSARs), it is often described as a 'black-box' method as it does not produce an explicit functional representation of the data (i.e. an algorithm). In GPR methods, the underlying function, $f(x)$, which produces the data will remain unknown and that the data are produced from an infinite set of functions with a Gaussian distribution in the function (chemical) space. The GPR model is fully characterised by its mean and covariance function[33] with the mean being considered the 'zero-everywhere' function. The mean of the GPR model is used as the predicted value (the output of the model), and the variance represents the error bars for this prediction. The molecular physicochemical descriptors of each molecule used in this study were selected as they are commonly used in similar modelling studies and are readily available without the need for specialist software.[19,21,28,30,34] Statistical performance measures for models are described below.

Analysis of data

Data from static and flow-through cell experiments were separated so that the performance of each group could be determined separately. In the first experiment, which used a leave-one-out methodology, the model is trained based on the flow-through data only and the predictions are obtained for flow-through data only. Similarly, in the second experiment, the model is based on the static diffusion cell data and the predictions are achieved only for static data. The effect of mixing data using both static and flow-through diffusion cells was then examined. Data from static diffusion cells and flow-through diffusion cells only were used to train each model and were also used to assess the performance of models which were trained with data from the other type of diffusion cell (i.e. models were based on training from static diffusion cells and then evaluated for flow-through cell data only, and *vice versa*).

Ten different training and test subsets from the mixed dataset were generated. As datasets for static and flow-through diffusion cell experiments are of different sizes (n = 91 and 53, respectively), the same number of static cell and flow-through cell data are included in each training set – 36 data were selected randomly for this purpose. The test set contains unequal measures of data from both static cell and flow-through cell datasets. This is repeated ten times in order to produce ten random training and test sets which are then trained separately and predictions obtained for their corresponding test sets. As the experiments have been repeated ten times, the results are reported as the mean value along with its standard deviation. Thus, the following experiments were conducted:

- *Experiment A*: Using data from static cell experiments only for training and test sets.
- *Experiment B*: Using data from flow-through cell experiments only for training and test sets.
- *Experiment C*: Using data from static cell experiments only for training and test sets from flow-through cell experiments only.
- *Experiment D*: Using data from flow-through cell experiments only for training and test sets from static cell experiments only.
- *Experiment E*: A mixed analysis of data collated from both datasets and used for training and test sets.

Part two – the effect of experimental temperature on model of skin permeation

In a separate experiment using the datasets published by Prapopoulou,[23] the effect of experimental temperature (T_{exp}) on model performance was evaluated. The numerical feature of experimental temperature (as T_{exp}) was added to the other descriptors examined in this study (see the section of Datasets). Initial studies showed that no benefit was obtained by including MR or SP into the analysis. The T_{exp} range used in this study was 22–45 °C (including studies which cite the skin surface temperature at 32 °C, rather than the temperature of the diffusion cells). All data that listed T_{exp} as 'not given' were omitted from the analysis.[22] The experimental results were compared before and after the inclusion of T_{exp} in the model. Analysis was conducted using GPR methods outlined above.

Performance measures

The performance of each model is determined, as in previous studies[21,28,30,35] by consideration of the correlation coefficient (r, CORR), the improvement over the Naïve model (ION) and the mean squared error (MSE). The training and test input target pairs are considered as (x_n^{trn}, y_n^{trn}) and (x_n^{tst}, y_n^{tst}) with a test input of x_n^{tst}. The MSE measures the average squared difference between a model prediction, denoted by $\hat{y}_n$, and the corresponding targets, y_n^{tst}. The normalised mean squared error, where the MSE is normalised by the variance of target values, is thus represented by the expression:

$$\mathrm{MSE} = \frac{1}{N_{tst}} \sum_{n=1}^{N_{tst}} (y_n^{tst} - \hat{y}_n)^2, \quad (1)$$

where N_{tst} denotes the number of target values in the test set.

The degree of improvement in the model, compared to the Naïve predictor (in the naive model for any input the prediction is always the same value, namely the mean of log K_p in the training set), is then represented by the expression:

$$\mathrm{ION} = \frac{\mathrm{MSE}_{naive} - \mathrm{MSE}}{\mathrm{MSE}_{naive}} \times 100\% \quad (2)$$

where MSE_{naive} represents the MSE of the naïve model.[21]

As well as reporting MSE and ION for each experiment, the correlation coefficient, r^2 or CORR, between targets and estimates is also reported to further characterise the quality of models. This is consistent with the use of r^2 in a range of predictive methods, including machine learning and quantitative structure–activity (or permeability) relationships (QSARs or QSPRs). This allows contextualisation of model quality with previous work in these fields. The aim of these performance measures is to obtain a model with low values of MSE and high values of both ION and CORR.

Results

Characterisation of the dataset

Figures 2–5 show the characterisation of the dataset. In Figure 2, the data are visualised in a matrix plot, comparing each individual molecular descriptor against each other. The data points have been shown in different signs and colours according to their labels: the static type is shown in red circles and the flow-through type by black crosses. In the diagonal (from top left to bottom right), the outline of grouped histograms is also plotted. It can be seen from this matrix that the distribution of the data from the static (Franz-type) diffusion cells is different to the distribution of the data from the flow-through cells across all the descriptors. The distribution of the molecular descriptors is shown for data from both static and flow-through cells in Figure 3 and indicates that the data from static diffusion cells have a similar mean value to data from flow-through cells for melting point, the Fedor's solubility parameter, log

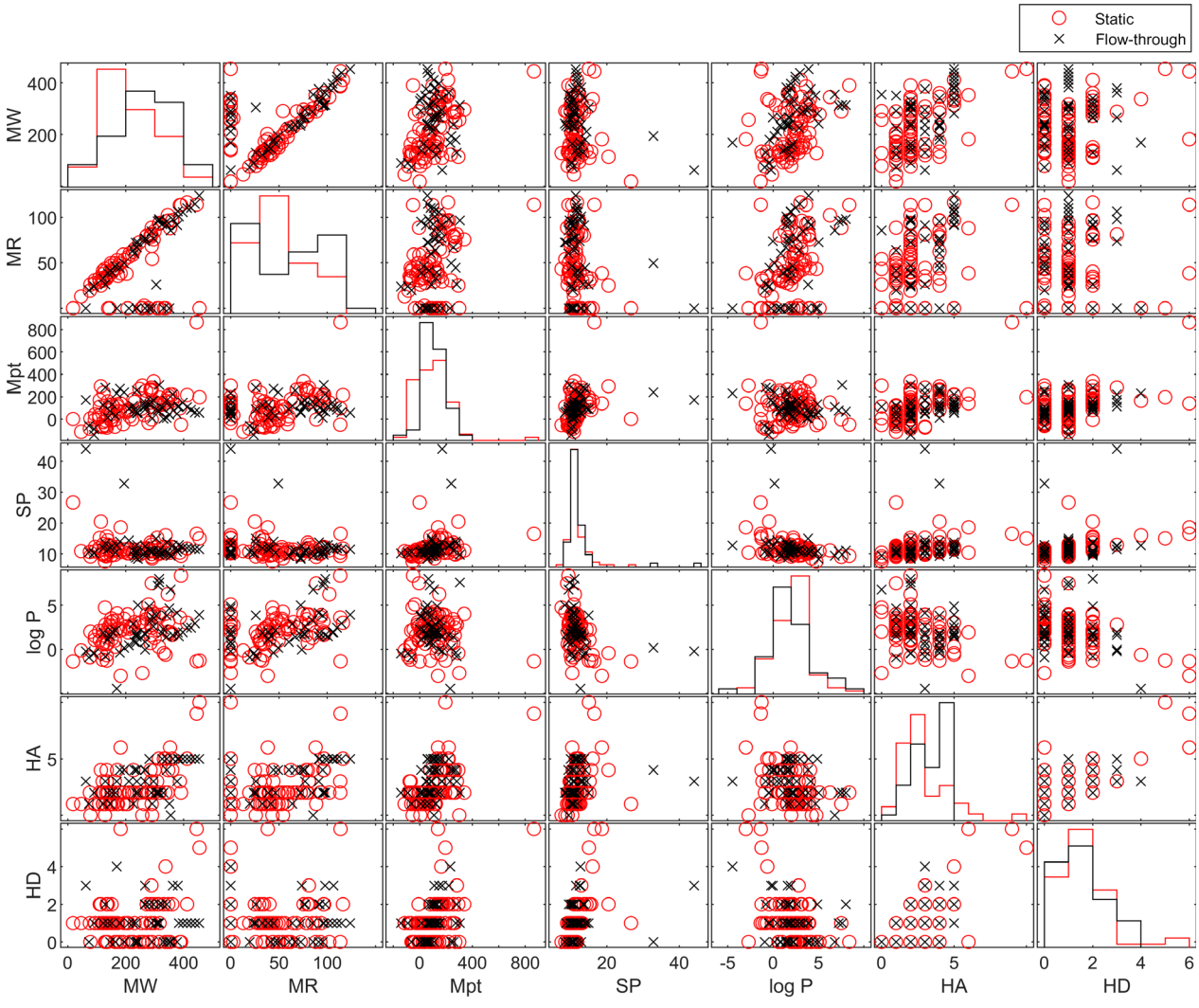

Figure 2 A matrix of the scatter plots of data used in this study, groups as being generated from either flow-through or static-type diffusion cells. [Colour figure can be viewed at wileyonlinelibrary.com]

P and the count of hydrogen bond donors, but that the means differ for molecular weight, molar refractivity and the count of hydrogen bond acceptors. Figure 4 shows the result of PCA analysis of the dataset. Plot (a) is a plot of the Eigenvectors, where each Eigenvalue indicates the amount of the variance within the data captured along the corresponding Eigenvector. The larger the Eigenvalue is, the more important the corresponding Eigenvector is, and the Eigenvector having the largest Eigenvalue is the first principal component. Thus, Figure 4a indicates that the first three components are relatively important in this PCA analysis, particularly the first two Eigenvectors which capture 67.97% of total variance. Plots (c) and (d) in Figure 4 also show that there is no linear relationship between log K_p and the compound features.

Figure 5 shows the results of the t-SNE analysis of the dataset. The basic aim of the t-SNE method is to minimise the divergence between the distribution that measures pairwise similarities of data in the original data space and the distribution that measures pairwise similarities of corresponding points in the low-dimensional space. Figure 5a shows the training errors, which converged after 100 iterations. Like the PCA plot shown in Figure 4, it is clear that the data from the two different types of diffusion cells are mingled together, implying that they occupy the same 'chemical space'. The complex and nonlinear structures shown in the data, as shown in Figure 5b–5d, indicate that that structures in the data is complex and that it is difficult to distinguish them by using these seven physicochemical features.

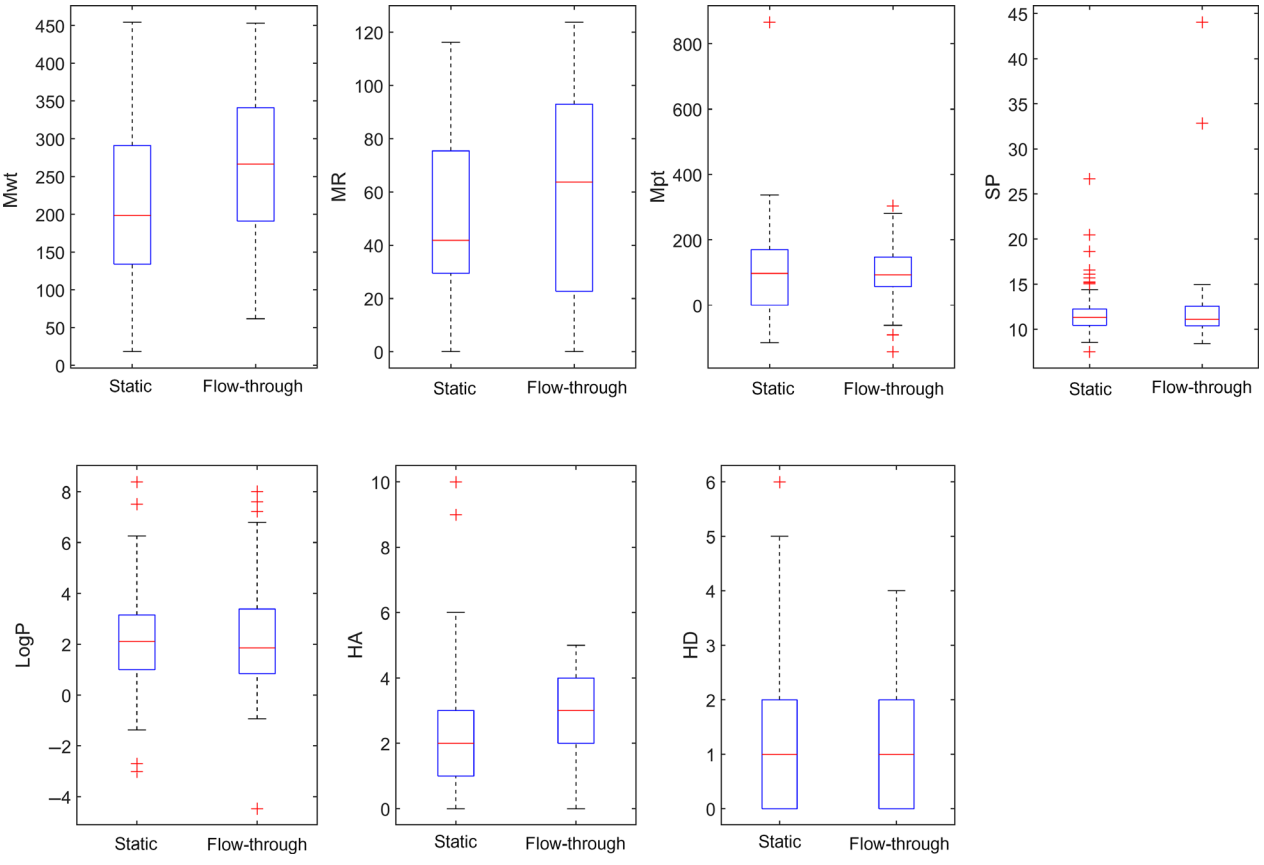

Figure 3 Comparison of box-plots (static vs flow-through diffusion cells) for each of the molecular descriptors used in this study. [Colour figure can be viewed at wileyonlinelibrary.com]

Effect of diffusion cell type on model quality

The results of experiments A and B are shown in Table 1, where it can be seen that the performance of the flow-through model (i.e., where permeation data from flow-through cell experiments only was used to train the model) is poor and almost the same as the naïve predictor. By comparison, the model using only static cell permeation data was substantially better (ION = 0.04_{FT} vs 0.42_{STAT}; correlation coefficient = 0.20_{FT} vs 0.66_{STAT}).

The data from flow-through and static cell experiments were then collated and analysed together as a single dataset in order to gauge the impact of datasets containing mixed static and flow-through diffusion cell data on model quality. The results of experiments C and D are shown in Table 2, and the average performance indicates that the best predictive models are always obtained when data from static diffusion cell experiments are used; when testing on the same data from static diffusion cell experiment, using the model trained by data from flow-through cell experiments, the value of ION is −7%, and the value of CORR is 0.19, whilst using the model trained by data from static cell experiment, the value of ION is 42% and the value of CORR is 0.66. Training models based on flow-through cell data only, and predicting the permeability of 'unseen' test data regardless of whether data from static or flow-through cells, resulted in poor models with a performance of 4% and 20% for ION and CORR, respectively (Table 1).

Data from the static and flow-through cell subsets were then collated together to determine the effect of mixing data from these different sources on model quality. The results of this analysis are shown in Table 3, which indicate that, although data from both sets are used to train the model, the prediction performance for flow-through diffusion cell data has not been improved. A very small increase in static diffusion cell data (Table 2; ION_{GP} 0.42, $CORR_{GP}$ 0.66 and Table 3; ION_{GP} 0.43, $CORR_{GP}$ 0.67) was observed. Thus, using data from Franz-type cell experiments to train a predictive model for flow-through diffusion cells, and vice versa (Table 3) resulted in poorly predictive models, suggesting a lack of comparability between the permeability data produced by both types of cells.

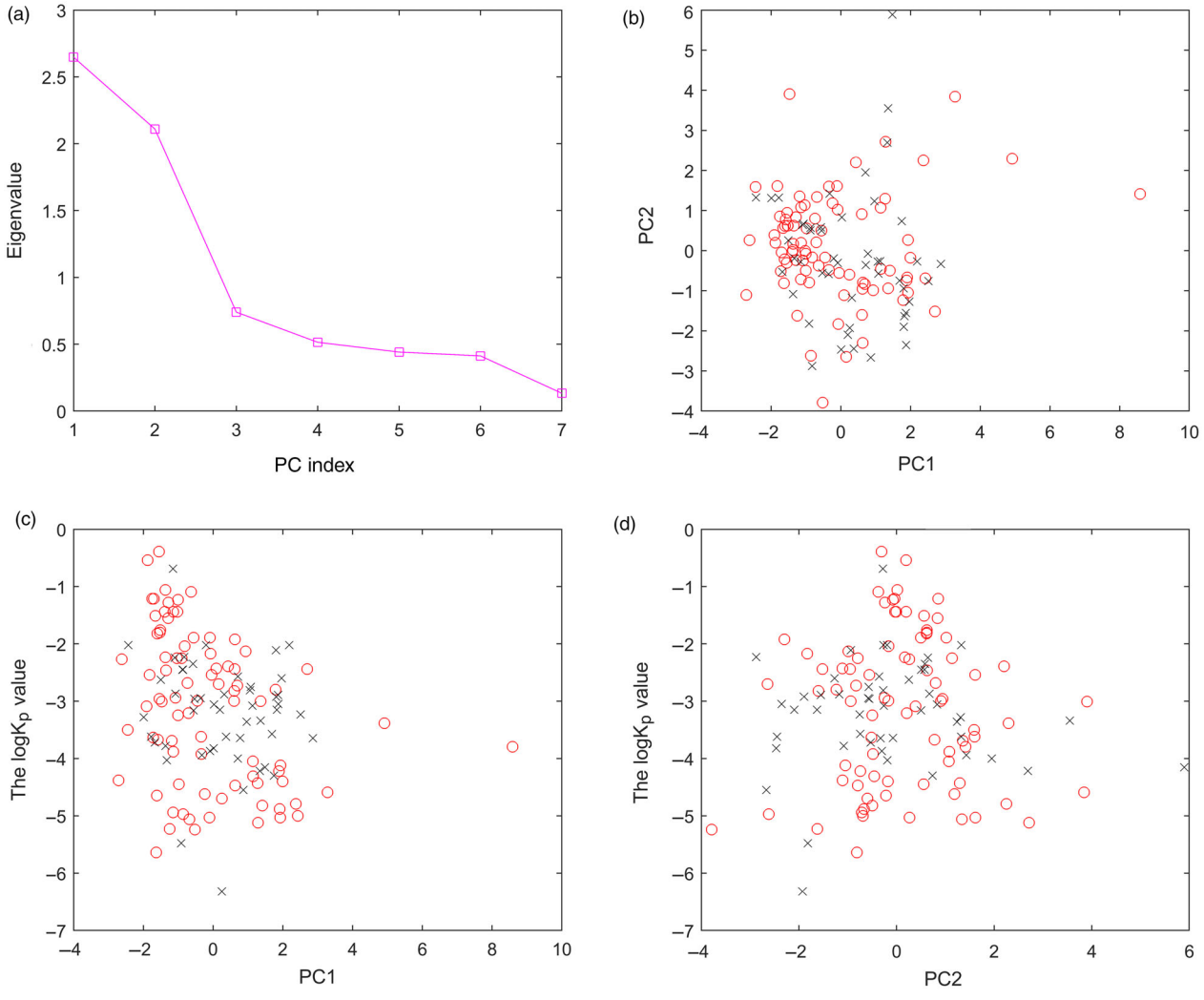

Figure 4 Principal component analysis (PCA) plot of the dataset used in this study, the data mapped to a low-dimensional space with a linear transformation, where (a) shows the Eigenvalue (variance) of each principal component; (b) shows a PCA plot of the first PC against the second the PC with data from static and flow-through cells mixed together. The compounds were plotted using the corresponding log K_p values against the first two principal components to represent the variation in the seven features of all chemical compounds, and this is shown in plot (c) for PC1 and plot (d) for PC2. Note that the red circle denotes the data obtained from static diffusion cells and the black crosses denote the data obtained from flow-through diffusion cells. [Colour figure can be viewed at wileyonlinelibrary.com]

Effect of experimental temperature (T_{exp}) on model performance

It was found that if the dataset contained a wider range of experimental temperatures, models with improved predictivity were obtained. For example, for the dataset with the widest range of temperatures (dataset F, Table 4), a 73% increase in ION was obtained. No improvement in model performance (as ION) was observed when the temperature range was restricted. This indicates that adding T_{exp} as a numerical feature to the data can be helpful in increasing model predictivity and that predictions are improved when the T_{exp} range in a dataset as a large as possible. Model performance was not substantially improved by adding additional physicochemical features to the analysis or by using SVM methods.[35]

Discussion

Diffusion cells for measurement of percutaneous absorption are comprised generically of two compartments separated by the membrane of interest. The donor

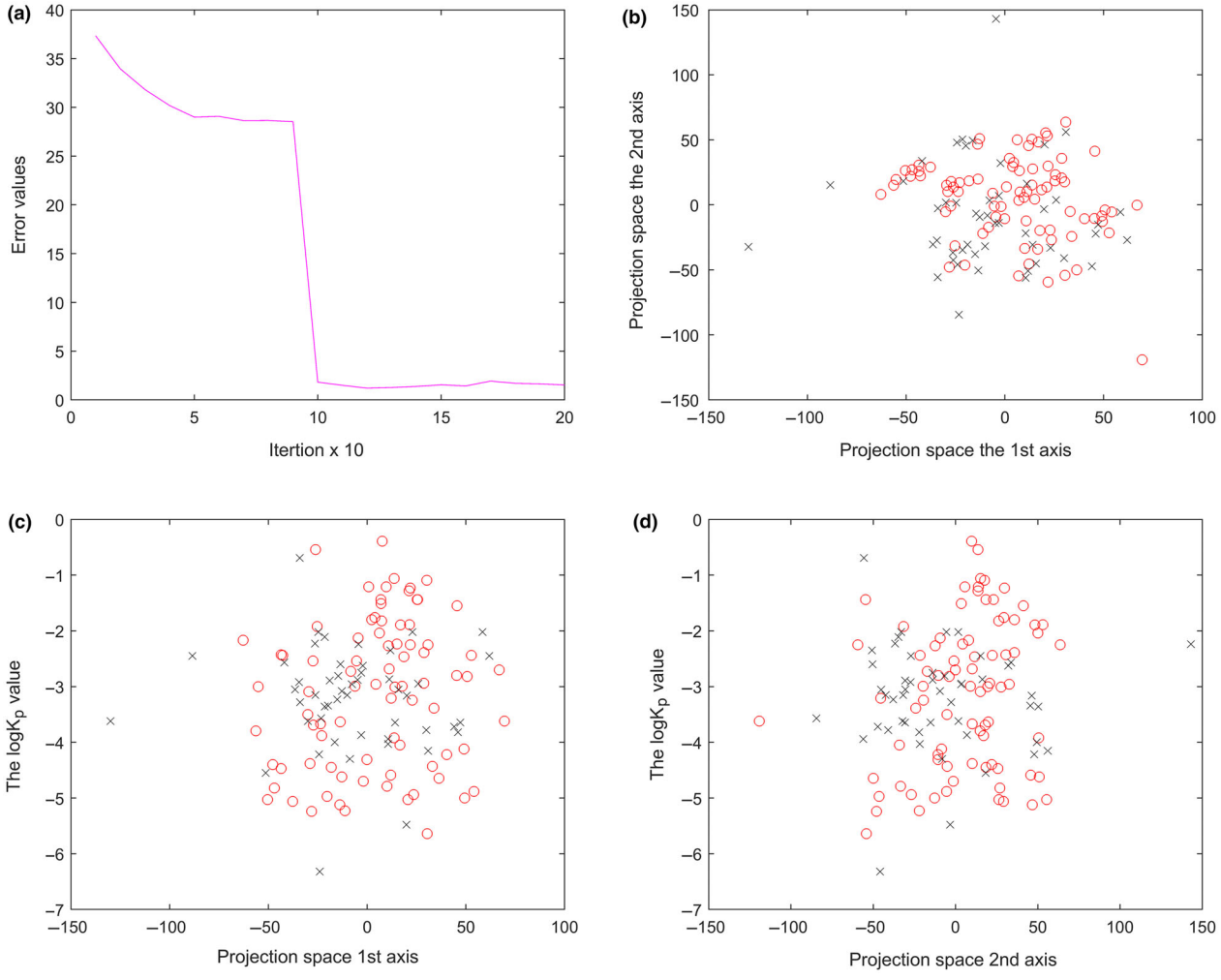

Figure 5 Results of t-SNE (t-distributed stochastic neighbour embedding) analysis, with (a) showing the training errors. Plot (a) shows the training errors. Plot (b) shows the projection plot using t-SNE. Plots (c) and (d) show the first and the second axis in the project space against the target (log K_p) values, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

compartment contains the penetrant, usually in a vehicle or other formulation, and it is collected from the receptor compartment after passing into and across the membrane of choice – in this context, the membrane is usually excised human or other mammalian skin or artificial membranes such as polydimethylsiloxane.

The two main types of diffusion cells for estimating the *in vitro* permeation of exogenous chemicals into and across skin are generally referred to as 'static', or Franz-type, diffusion cells[36] or flow-through, or in-line or 'Bronaugh-type' diffusion cells[37,38] (Figure 1). The static cell is usually maintained in an upright position, with the receptor compartment (usually 2–20 ml in volume, compared to a flow rate in flow-through cells of 1–2 ml/min[39]) continually stirred to ensure even distribution of the permeant and is kept at a particular temperature which reflects either the *in vivo* situation or the needs of particular penetrants.[40] Whilst static cells find greater utility in the field – ostensibly due to convenience and cost – a number of studies indicate that there are no differences between permeability data obtained using either method.[41–44] In addition, the inherently variable nature of the skin membrane is considered in such experiments to exert a significant effect which may make any comparisons difficult or irrelevant. Official guidelines for *in vitro* diffusion cell studies, such as OECD 428, indicate that the use of either type of cell is acceptable.[45] In general it is perceived that both static and flow-through cells are similar in their production of permeability data and that the overall experimental design – notably the maintenance of skin conditions or the occlusivity of the

Table 1 Results of experiments A and B: prediction performances for static and flow-through diffusion cells used to assess the models

	Flow-through cell		Static cell	
	Mean	STD	Mean	STD
MSE_{GP}	0.84	0.12	0.98	0.13
ION_{GP}	0.04	0.04	0.42	0.09
$MSE_{NaiveGP}$	0.87	0.13	1.68	0.09
$CORR_{GP}$	0.20	0.13	0.66	0.06

Table 2 Results of experiments C and D: prediction performances for a single dataset (with collated data from flow-through and static diffusion cell experiments) used to assess the models

	Flow-through cell		Static cell	
	Mean	STD	Mean	STD
MSE_{GP}	0.93	0.23	0.96	0.09
ION_{GP}	−0.07	0.09	0.43	0.05
$MSE_{NaiveGP}$	0.86	0.14	1.70	0.09
$CORR_{GP}$	0.19	0.16	0.67	0.05

donor compartment – is more important in the generation of valid and reliable data from *in vitro* experiments.[43]

However, the results of this study indicate that the quality of the model is directly affected by the inclusion of data from flow-through experiments which reduces overall model quality and predictive power. Models based solely on flow-through data offer poor predictions of skin permeability, compared to models based on data derived from static diffusion cell experiments which resulted in comparatively highly predictive models. Thus, in order to optimise the model quality, data from only static, Franz-type, experiments should be used to construct the model and that data from flow-through studies should not be used for this purpose as it yields by itself very poor models and also reduces the predictive accuracy of models when mixed with data from static diffusion cell experiments.

Table 3 Results of experiments E: performance measures for training flow-through or static cell models with data from the other experiments

	Static cell data to train a predictive model for a flow-through cell model		Flow-through cell data to train a predictive model for a static cell model	
	Mean	STD	Mean	STD
MSE_{GP}	1.17	0.32	1.64	0.06
ION_{GP}	−0.35	0.22	0.05	0.04
$MSE_{NaiveGP}$	0.85	0.24	1.73	0.16
$CORR_{GP}$	0.07	0.14	0.19	0.10

Despite the overall validation of flow-through diffusion cells proposed by Addicks *et al*.,[41] they determined that, for a series of alkyl *p*-aminobenzoates diffused through a PDMS membrane, certain specific differences were observed when compared to static diffusion cells. Permeability coefficients for each of the compounds were obtained from static and flow-through diffusion cells and indicated that whilst the methyl and ethyl compounds were not significantly different, the permeability coefficient for the propyl ester, at a flow rate of 24 ml/h, was significantly higher than for either the 12 ml/h flow rate or for the Franz-type diffusion cell. Permeability coefficients at both flow rates in the flow-through diffusion cell were found to be significantly higher than those obtained with the Franz-type diffusion cell for both butyl and pentyl esters ($P < 0.05$). In exploring the flow rates further it was found that results were more erratic, with large standard deviations as the flow rate was increased to, for example, 60 ml/h. They proposed that this was due to turbulent flow of the perfusant, which resulted in eddies in the current, and fluid channels in the cell which resulted in the formation of a large hydrodynamic layer which may affect perfusant solubility. They proposed that an optimum flow rate was required in flow-through cells which maintained sink conditions (which in their studies they found to be no more than 7% of the donor concentration at a flow rate of 12 ml/h, lower than the often-cited threshold of 10% by, e.g. Barry[46]).

It is important to consider in this context some of the data used to construct mathematical models of skin permeability and their experimental conditions – and to note that such data is usually derived for purposes other than the development of mathematical models. In considering previous studies[11,12,44,47] whose data has been harvested in order to populate datasets from which various algorithms of skin permeability have been derived, the flow rates in flow-through cell experiments is different in all these studies, ranging from 1.1 to 5 ml/h, which is both variable and outside the optimum proposed by Addicks *et al*.[41] Building on their comments on flow rate and maintenance of skin conditions, it might be proposed that the flow rates in these studies are outside the ideal range to ensure full solubility.

The wider point is not to consider the *specifics* of these different studies, which may have particular requirements which suit the nature of the chemicals of interest in each study. More important is the perception that data from static and flow-through cell experiments is interchangeable and can be collated into larger databases which are then used to develop predictive models. The design of datasets and the volume of data required to construct a robust model have been considered previously.[22] The findings of this study add to[22] by indicating that a closer inspection

Table 4 ION performance with and without T_{exp} added to the five physicochemical descriptors used

Using five physicochemical descriptors	Dataset A	Dataset B	Dataset C	Dataset D	Dataset E	Dataset F
Range of experimental temperatures in each dataset (°C)	Temperature listed as either 37 °C or 37 °C with a skin surface temperature of 32 °C				22–45	22–45
Size of dataset (original/refined [22])	11/9	42/25	38/21	99/57	92/51	148/86
ION (mean value, without including T_{exp} in the analysis)	0.19	−0.03	0.38	0.33	0.00	0.37
ION (mean value, including T_{exp} in the analysis)	0.19	−0.03	0.38	0.33	0.01	0.64

Data was refined as in [22] by, for example, removing ambiguous data or values, which are listed as 'greater than' or 'less than' a fixed value, rather than a discrete number.

of the experimental protocols for the data is required before data can be considered for use in datasets. In this context, it is also important to consider that the design of diffusion cells is highly variable in size and that the maintenance of sink conditions is vital. Flow-through cells vary in terms of the materials used (e.g. some flow-through cells are manufactured from Teflon™,[37] whilst the vast majority of Franz-type cells are manufactured from glass) flow rate, possible issues of sample loss (if, for example, chemicals of interest are liable to be absorbed into the tubing used to perfuse the receptor compartment or if the temperature of significant amounts of tubing is not carefully controlled) and the gap between perfusion and sample collection. Whilst *in vitro* permeation data are normally interpreted in this context and accommodation in interpretation of data follows, this is not necessarily the same for models derived from collated literature data. There are echoes of this outcome in previous parts of the literature in this field. For example, the re-analysis of steroid permeability data used in Flynn's dataset[1,48] and its subsequent QSAR remodelling[34] indicated that steroids were no longer to be considered outliers which permeated the skin by a different mechanism of action, but that that conclusion had been made based on erroneous data which is still considered in some recent models of skin permeation.[49]

The effect of experimental temperature on model quality was also considered. In this study, it was determined that, when T_{exp} was considered as a descriptor of permeability in the same manner as the key physicochemical descriptors of a molecule, it significantly influenced the nature of the resultant model; specifically, using a wider range of experimental temperatures improved model quality in terms of its improvement over the naïve model (Table 4). It is interesting to note that this effect was more pronounced compared to adding additional physicochemical descriptors to the model or when using different machine learning methods, such as Support Vector Machines.[35] Whilst this result appears anomalous and possibly contradictory, it does echo other findings for permeation across a PDMS membrane in which T_{exp} was found to be significant [50]. Further, it supports the analysis of diffusion cell type in that it indicates how experimental conditions can influence the construction of models, and underpins our recommendation that such descriptors of the experiment from which the data is derived, and not just the physicochemical descriptors of chemicals, be included in the construction of models derived from literature data.

Conclusions

The experimental conditions examined in this study – diffusion cell type and T_{exp} – are shown to have significant effects on the quality of models derived from these different studies and can significantly affect the outputs (predictions of permeability and elucidation of mechanism insight) of these models. Separating the data based on diffusion cell type – as either static, Franz-type or flow-through, Bronaugh-type diffusion cells – shows that the best predictive models are always obtained when static diffusion cell skin permeability data is predicted compared to models constructed from flow-through cell experiments. These results are obtained regardless of whether data from static or flow-through cell experiments are used to train models. However, training models based on flow-through cell data only which is used to train 'unseen' test data (randomly taken from either diffusion cell dataset) resulted in models of poor statistical quality and limited predictive power. Conversely, when a wide range of experimental temperatures is used the performance of models improves substantially.

This study indicates that model quality is also influenced significantly by experimental factors such as diffusion cell type and experimental temperature. Models produced from static, Franz-type, diffusion cell studies, which resulted in highly predictive models. It is therefore clearly recommended from the findings of this study that, in order to optimise the predictive power of a mathematical model of skin permeation, data from static cell experiments only should be used to construct the model and that relevant experimental conditions be incorporated into the model where relevant.

Declarations

Conflict of interests

The Authors have no conflict of interests to report.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Acknowledgements

The authors would like to thank the University of Hertfordshire and Keele University for supporting this study.

References

1. Flynn GL. Physicochemical determinants of skin absorption. In: Gerrity TR, Henry CJ, eds. *Principles of Route-to-Route Extrapolation for Risk Assessment*. New York, NY: Elsevier, 1990: 93–127.
2. Anderson BD *et al.* Heterogeneity effects on permeability-partition coefficient relationships in human stratum corneum. *Pharm Res* 1988; 5: 566–573.
3. Chowhan Z, Pritchard R. Effect of surfactants on percutaneous absorption of naproxen. 1. Comparisons of rabbit, rat and human excised skin. *J Pharm Sci* 1978; 67: 1272–1274.
4. Dutkiewicz T, Tyras H. Skin absorption of toluene, styrene and xylene by man. *Br J Ind Med* 1968; 25: 243.
5. Jolicoeur LM *et al.* Etorphine is an opiate analgesic physicochemically suited to transdermal delivery. *Pharm Res* 1992; 9: 963–965.
6. Dutkiewicz T, Tyras H. A study of the skin absorption of ethylbenzene in man. *Br J Ind Med* 1967; 24: 330–332.
7. Hadgraft J, Ridout G. Development of model membranes for percutaneous absorption measurements. 1. Isopropyl myristate. *Int J Pharm* 1987; 39: 149–156.
8. Michaelis AS *et al.* Drug permeation through human skin: theory and in vitro experimental measurement. *AIChE* 1975; 21: 985–996.
9. Raykar PV *et al.* The role of protein and lipid domains in the uptake of solutes by human stratum corneum. *Pharm Res* 1988; 5: 140–150.
10. Roy SD, Flynn GL. Transdermal delivery of narcotic analgesics – comparative permeabilities of narcotic analgesics through human cadaver skin. *Pharm Res* 1989; 6: 825–832.
11. Roberts MS *et al.* Permeability of human epidermis to phenolic compounds. *J Pharm Pharmacol* 1977; 29: 677–683.
12. Roberts MS *et al.* Percutaneous absorption of phenolic compounds – mechamism of diffusion across stratum corneum. *J Pharm Pharmacol* 1978; 30: 486–490.
13. Roy SD, Flynn GL. Transdermal delivery of narcotic analgesics – pH, anatomical and subject influences on cutaneous permeability of fentanyl and sufentanyl. *Pharm Res* 1990; 7: 842–847.
14. Shaw JE, Chandrasekaran SK. Controlled delivery of drugs for systemic action. *Drug Met Rev* 1978; 8: 223–233.
15. Scheuplein RJ, Blank IH. Permeability of the skin. *Physiol Rev* 1971; 51: 702–747.
16. Scheuplein RJ. Mechanism of percutaneous absorption. I. Routes of penetration and the influence of solubility. *J Invest Dermatol* 1965; 45: 334–346.
17. Scheuplein RJ *et al.* Percutaneous absorption of steroids. *J Invest Dermatol* 1969; 52: 63–70.
18. Moss GP *et al.* Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Tox In Vitro* 2002; 16: 299–317.
19. Moss GP *et al. Predictive Methods in Percutaneous Absorption*. Heidelberg: Springer AG, 2015.
20. Hewitt M *et al.* QSAR and machine learning analysis of a PDMS dataset. In: Brain KR, Walters KA, eds. *Proceedings of the 14th Perspectives in Percutaneous Penetration Conference*, Vol. 14. Cardiff: STS Publishing, 2014: 87. ISBN 978-0948917-48-6.
21. Moss GP *et al.* The application of Gaussian processes to the prediction of percutaneous absorption. *J Pharm Pharmacol* 2009; 61: 1147–1153.
22. Ashrafi P *et al.* Model fitting for small skin permeability data sets: hyperparameter optimisation in Gaussian Process Regression. *J Pharm Pharmacol* 2018; 70: 361–373.
23. Prapopoulou M. The development of a computation/mathematical model to predict drug absorption across the skin. PhD Thesis, King's College London, 2012. Available at: https://kclpure.kcl.ac.uk/portal/. Accessed 24th July 2019.
24. Wilschut A *et al.* Estimating skin permeation: the validation of five mathematical models. *Chemosphere* 1995; 30: 1275–1296.
25. Patel H *et al.* Quantitative structure-activity relationshops (QSARs) for prediction of skin permeation of exogenous chemicals. *Chemosphere* 2002; 48: 603–613.
26. Moss GP *et al.* Design, synthesis and characterisation of captopril prodrugs for enhanced percutaneous absorption. *J Pharm Pharmacol* 2006; 58: 167–177.
27. Soyei S, Williams F. The EDETOX Database. Available at: https://research.ncl.ac.uk/edetox/theedetoxdatabase/. Accessed 24th July 2019.
28. Sun Y *et al.* The application of stochastic machine learning methods in the prediction of skin penetration. *App Soft Comp* 2011; 11: 2367–2375.
29. Fedors RF. A method for estimating both the solubility parameters and molar volumes of liquids. *Poly Eng Sci* 1974; 14: 147–154.
30. 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.
31. van der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. *J Mach Learn Res* 2008; 9: 2579–2605.
32. Ashrafi P *et al.* The application of Machine Learning to the modelling of percutaneous absorption: an overview

and guide. *SAR QSAR Environ Res* 2015; 26: 181–204.
33. Rasmussen CE, Williams CKI. *Gaussian Processes for Machine Learning*. Cambridge: The MIT Press, 2006.
34. Moss GP, Cronin MTD. Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data. *Int J Pharm* 2002; 238: 105–109.
35. Moss GP *et al*. The application of discriminant analysis and Machine Learning methods as tools to identify and classify compounds with potential as transdermal enhancers. *Eur J Pharm Sci* 2012; 45: 116–127.
36. Franz TJ. On the relevance of in-vitro data. *J Invest Dermatol* 1975; 64: 190–195.
37. Bronaugh RL, Stewart RF. Methods for in-vitro percutaneous absorption studies. IV: the flow-through diffusion cell. *J Pharm Sci* 1985; 74: 64–67.
38. Bronaugh RL *et al*. Methods for in-vitro percutaneous absorption studies. VII: use of excised human skin. *J Pharm Sci* 1986; 75: 1094–1097.
39. Akhtar SA *et al*. An automated diffusion apparatus for studying skin penetration. *Int J Pharm* 1984; 21: 17–26.
40. Williams AC. *Transdermal and Topical Drug Delivery*. London: The Pharmaceutical Press, 2003.
41. Addicks WJ *et al*. Validation of a flow-through diffusion cell for use in transdermal research. *Pharm Res* 1987; 4: 337–341.
42. Chatteraj SC, Kanfer I. Release of acyclovir from semi-solid dosage forms: a semi-automated procedure using a simple plexiglass flow-through cell. *Int J Pharm* 1995; 125: 215–222.
43. Clowes HM *et al*. Skin absorption: flow-through or static diffusion cells? *Tox In Vitro* 1994; 8: 827–830.
44. Cordoba-Diaz M *et al*. Validation protocol of an automated in-line flow-through diffusion equipment for in-vitro permeation studies. *J Cont Rel* 2000; 69: 357–367.
45. OECD Guideline for the Testing of Chemicals. *428: Skin Absorption: In-vitro Method*. Paris: OEC, 2004.
46. Barry BW. *Dermatological Formulations: Percutaneous Absorption*. New York, NY: Marcel Dekker, 1983.
47. Beckley-Kartey SA *et al*. Comparative in-vitro skin absorption and metabolism of coumarin (1,2-benzopyrone) in human, rat and mouse. *Toxicol Appl Pharmacol* 1997; 145: 34–42.
48. Johnson ME *et al*. Permeation of steroids through human skin. *J Pharm Sci* 1995; 84: 1144–1146.
49. Keurentjes AJ, Maibach HI. Percutaneous penetration of drugs applied in transdermal delivery systems: an in-vivo based approach for evaluating computer generated penetration models. *Regul Toxicol Pharmacol* 2019; 108: 104428.
50. Moss GP *et al*. NARMAX models in the prediction of penetration across skin and polydimethylsiloxane membranes. In: Brain KR, Chilcott R, eds. *Advances in the Dermatological Sciences*. Cambridge: Royal Society of Chemistry, 2013: 384–388.