

# Predicting Drug Absorption Rates Through Human Skin

Yi Sun, Lun Tak Lam, , Gary Moss, Maria Prapopoulou, Rod Adams, Neil Davey, David Gray, Mark Brown<sup>1</sup>

**Abstract**—Predicting the rate at which a substance will pass through human skin and into the bloodstream is a problem of current interest. We use Gaussian Process modeling to train a set of predictors using every combination of six molecular features. We find that only three of the features are needed for our best predictor. This result could be useful in the further analysis of skin permeability.

## I. INTRODUCTION

In recent years there has been increasing attention paid to the problem of predicting the rate at which a substance will pass through human skin [1-6]. Interest in this issue has been driven by the pharmaceutical industry, where it is important to be able to predict the rate at which a drug will pass through the skin and into the bloodstream. Such knowledge enables skin patches to be designed to administer drugs at appropriate dosages. The issue is also of interest to the cosmetics industry where the concern is normally to reduce dermal uptake. For example it is not desirable for a sun block to pass readily into the bloodstream. Lastly it is important to know how much of an industrial or household chemical is likely to pass through skin should accidental contact arise.

Of course the simplest way to predict the rate at which a chemical will be absorbed is by direct measurement. This process, however, is difficult, expensive and time consuming. So attempts have been made to use known absorption rates as the basis for extrapolation to new compounds.

The first attempts to do this took two easily measured features of a compound, the *lipophilicity* and the *molecular weight*, and then tried to fit a linear regression to the data, with the absorption rate as dependent variable. Subsequently this simple linear model has been refined and modified. More recently non-linear models such as neural nets [2] and *k*-nearest neighbours have been tried. However the number of available data points is not large and this encouraged us to investigate how well a Gaussian Process model (*GP*) would perform on this data. We also look at a wider range of molecular features than just the two aforementioned measures. The novel contribution described here is that we investigate every combination of six molecular features to find the set that provides the best overall prediction.

## II. BACKGROUND

The outer layer of human skin, the *stratum corneum*, is the primary barrier to absorption. It is a thin layer, 15 – 30 $\mu$ m, deep and it consists of dead keratin cells in a lipid domain, the so called bricks and mortar model. For a chemical to pass through the dermis, in any quantity, it must be desolvable in lipids (have high lipophilicity) and small, so that it can diffuse through the barrier. The lipophilicity is measured as the *partition coefficient*, *P*. This quantifies the extent to which the chemical prefers to be in the particular lipid *octanol* rather than water. It is straightforward to measure this. The chemical under investigation is dissolved in a equal mixture of octanol and water. The mixture is then allowed to separate – it takes about 24 hours, and the relative quantities of the dissolved chemical, in each domain, are then measured. As this value varies over a very large range,  $10^{-7}$  to  $10^{+7}$  it is usual to use the logarithm of *P*.

Crudely speaking big molecules find it harder to get through the stratum corneum than do small molecules, so the molecular weight *MW*, is also commonly used to indicate permeability.

To measure the actual permeability of skin, a chemical is placed on the outer surface of the skin and the rate at which it is absorbed is measured. This rate is concentration corrected by dividing by the concentration of the chemical on the external surface of the skin, as the higher the concentration the higher will be the absorption rate. This gives the *permeability coefficient*,  $K_p$ , of the chemical. Again this has a very large range so  $\log K_p$  is normally reported.

In 1992 the first attempt was made at predicting permeability from *P* and *MW*. Potts and Guy [6] found the best linear fit for a data set of 93 compounds. Their actual equation is:

$$\log(K_p) = 0.71 \log(P) - 0.006 MW - 6.3$$

Such linear fits are known as Quantitative Structure-Permeability Relations (*QSPRs*) to the pharmacologists. The regression was not very successful with a coefficient of determination of 0.67 (1 is perfect, 0 is bad). This suggested that the relationship is non-linear and may involve other physical/chemical features of the molecule. So in the intervening years several researchers have proposed other linear models using a variety of molecular descriptors. However none of these has been particularly successful across a wide range of different molecular types. There has also been a small amount of work looking at non-linear models [2], although none of these have used

---

<sup>1</sup> The authors are at the University of Hertfordshire, Hatfield, UK. Corresponding author Yi Sun, y.2.sun@herts.ac.uk

Gaussian Process modeling.

### III. THE DATA

The data set used here consists of the original Potts and Guy data supplemented by a variety of data from other sources. It contains data on 142 molecules and a detailed description can be found in [5]. As well as molecular weight and lipophilicity we also have the four following additional molecular features that are thought to be important in determining a molecules ability to pass through skin:

- The melting point, *MPt*
- The solubility parameter *SP*. This measures how soluble the chemical is in the stratum corneum
- *HA* and *HD*, which are counts of the number of hydrogen bonding acceptor (*HA*) and donor groups (*HD*), respectively, that can be found on a molecule.

So in summary our data set is 142 6-ary vectors each of which has a measured value of its skin permeability coefficient  $\log K_p$ .

### IV. GAUSSIAN PROCESS MODELLING

*Gaussian process* (GP) modelling [7] is a non-parametric method, which does not produce an explicit functional representation of the data. Here it is assumed that the underlying function,  $f(\mathbf{x})$ , that produces the outputs will remain unknown, but that the data is produced from a (infinite) set of functions, with a Gaussian distribution in the function space. A multivariate Gaussian distribution is characterised by its mean and covariance matrix. Similarly a *Gaussian process* is completely characterised by its mean and covariance *function*. For simplicity, we usually consider the mean function to be the zero everywhere function. The covariance function,  $k(\mathbf{x}_i, \mathbf{x}_j)$ , is crucial to GP modelling.

It expresses the expected correlation between the values of  $f(\mathbf{x})$  at  $\mathbf{x}_i, \mathbf{x}_j$ . In other words, it defines the similarity of data points. It is normal to incorporate noise into the model, as the original data will probably be noisy. Here, we use the squared exponential covariance function, which incorporates noise, as follows:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)\right) + \delta_{ij} \sigma_n^2,$$

where  $\mathbf{M} = l^{-2} \mathbf{I}$ ,  $l$  is the *characteristic length-scale*,  $\sigma_f$  is *signal variance*, is  $\sigma_n$  *noise variance*, and  $\delta_{ij}$  is one if  $i = j$  and zero otherwise. It is also possible to have different length scales for each feature of the input. In this case the diagonal of  $\mathbf{M}$  is replaced with the vector  $[l_1^{-2}, l_2^{-2}, \dots, l_N^{-2}]$ ,

where  $l_i$  is the length scale of feature  $i$ . Note that the larger the value of  $l_i$  the less important is that feature. The central exponential gives a function that falls away rapidly as the distance between the two data points increases.

Using this covariance function we prepare to predict the output value  $y_*$  for a novel input vector  $\mathbf{x}_*$ . We first compute the covariance between all possible pairs of data points. We denote  $\mathbf{K}$  as the  $N$  by  $N$  matrix of covariances of points in the training set, that is  $\mathbf{K}(i, j) = k(\mathbf{x}_i, \mathbf{x}_j)$ ,  $\mathbf{K}_*$  as the  $N$ -ary vector of covariances of the  $N$  training points with the new input, so  $\mathbf{K}_*(i) = k(\mathbf{x}_i, \mathbf{x}_*)$  and  $\mathbf{K}_{**}$  is the covariance of  $\mathbf{x}_*$  with itself,  $k(\mathbf{x}_*, \mathbf{x}_*)$ .

To make the prediction,  $y_*$  at the new input  $\mathbf{x}_*$ , we need to compute the conditional distribution  $p(y_* | y_1, y_2, \dots, y_N)$  given the already observed:  $y_1, y_2, \dots, y_N$ .

Since the model is a Gaussian process, this distribution is also a Gaussian and is completely defined by its mean and variance. The mean of the distribution can be shown to be:

$$\langle y_* \rangle = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y}$$

and the variance is:

$$\text{var}(y_*) = \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T$$

So the predicted value will be  $\langle y_* \rangle$  and the variance gives a measure of our confidence in the prediction.

The three hyperparameters of the model, the *signal variance*, the *noise variance* and the *length scale vector*, can be estimated by expectation maximization on the training set. So under the assumption that the training data was generated by a GP, a search through the hyperparameter space is undertaken for the parameters that make the observed data most likely. Since the length scale vector can be found in this way a method of Automatic Relevance Detection is readily available. In other words, if a feature is found to have a large lengthscale it will contribute very little to the covariance and thus has low relevance.

### V. PERFORMANCE MEASURES

The standard linear regressors often used in this field do not perform very well. So for a baseline comparison we use a *naïve* predictor, that always predicts the same value, namely the mean value of permeability over the complete training set. So we first calculate the mean square error (*MSE*) of our predictor and report the *Improvement On the Naïve* (*ION*) predictor as:  $ION = \frac{MSE_{naive} - MSE}{MSE_{naive}}$ . We also

report the normalized mean square error, where the *MSE* is divided by the variance of  $\log K_p$  in the test set, and finally we give the correlation coefficient of the predictions with

the measured values, over the test set.

#### IV. THE EXPERIMENTS

Since the data set here is very small we divided it into a training set of 75% of the data (107 molecules) and training set of the remaining 25% (35 molecules). For each experiment reported we validate the training/test split by doing 10 independent runs with different random training/test splits. The results for the models are therefore averages over these 10 runs.

##### Experiment 1 – QSPR Method

In this experiment we compared two trainable models with the best performing of the linear predictors, the QSPRs, used in the field. On our data the model of Lao et al. was the best performing and its form is:

$$\log K_p = 0.5752 \log P - 0.004475 MW - 2.64368$$

For the comparison to be fair the trainable classifiers were only given the two features used by the QSPR, namely lipophilicity and molecular weight. As stated earlier the non-linear model used was a Gaussian Process regressor with hyper parameters set from the training set. In order to see how well the QSPR did against a linear predictor optimised for this specific data set a simple, single layer, linear neural network (NN) was also trained on the data. The results are shown in Table 1.

TABLE 1

| Model | ION (%)        | NMSE        | Correlation Coefficient |
|-------|----------------|-------------|-------------------------|
| QSPR  | -53.25 ± 36.24 | 1.56 ± 0.35 | 0.24 ± 0.11             |
| NN    | 9.83 ± 11.1    | 0.93 ± 0.17 | 0.34 ± 0.17             |
| GPR   | 22.89 ± 10.62  | 0.79 ± 0.17 | 0.49 ± 0.11             |

It is clear that the QSPR does not do at all well on this data set; it predicts a poorer value than the naïve predictor. The reason for this is that this regressor was designed for a different set of molecules (with some overlap) than those used here. It has not generalised well. The single layer network (NN), also a linear model, has done better and does improve on the baseline. The Gaussian Process model is clearly the best and this confirms our earlier results [5] that suggested that skin permeability prediction needs a non-linear model. Figure 1 illustrates the non-linearity of this problem. The skin permeability, as predicted by the GP, is plotted against  $\log P$  (vertical axis) and  $MW$  (horizontal axis).

##### Experiment 2 – Feature Selection

Each molecule can be represented by up to 6 features, and due to the small nature of our data we were able to complete a full evaluation of all combinations of these features. In other words we formed data sets in which each molecule is represented by every possible pair of features

(15 different sets), by each triple, and so, on until all six features were used.

QuickTime™ and a decompressor are needed to see this picture.

Fig. 1: A contour map of the predicted skin permeability against  $MW$  and  $\log P$ . The non linearity of the prediction is clear.

This gives a total of 57 different data sets. Both of our trainable models were trained using these 57 data sets. First the 6 best performing GP models are shown in Table 2.

TABLE 2  
GP PREDICTION WITH THE BEST PERFORMING FEATURES

| Combination of features | NMSE ± SD   |
|-------------------------|-------------|
| MPt, Log P, HD          | 0.64 ± 0.13 |
| MW, MPt, Log P, HD      | 0.65 ± 0.15 |
| MW, MPt, SP, Log P, HD  | 0.65 ± 0.14 |
| MW, SP, Log P, HD       | 0.67 ± 0.18 |
| MPt, SP, Log P, HD      | 0.67 ± 0.12 |
| MW, Log P, HD           | 0.68 ± 0.19 |

It can be seen that all the best performing models used  $\log P$ , the lipophilicity and  $HD$ , the number of hydrogen donor groups. As well as these two the melting point,  $MPt$ , and/or molecular weight,  $MW$ , were also included. In fact from the modeling point of view melting point and molecular weight are interchangeable.

As stated earlier it is possible to infer the length scale vector for the GP model from the training data. Table 3 shows the length scales of each of the six inputs in a variety of different models

It can be seen that for most variable the length scales are of the same order of magnitude. However the solubility parameter  $SP$ , consistently has large length scales. This suggests that for this data  $SP$  does not provide useful information for predicting permeability. It is notable that whilst some features, such as  $\log P$ , have generally low lengthscales the importance of the features varies with the context in which they are used. For example in the last model ( $MW$ ,  $MPt$ ,  $HA$ ,  $HD$ ) molecular weight is very important whereas in the second model ( $MW$ ,  $MPt$ ,  $\log P$ ,  $HD$ ) it is not so.

TABLE 3  
LENGTH SCALES FOR A VARIETY OF FEATURE SETS

**Length scale**

| <b>Combination of features</b> | <b>MW</b> | <b>MPt</b> | <b>SP</b> | <b>Log P</b> | <b>HA</b> | <b>HD</b> | <b>Features significance Ranking</b> |
|--------------------------------|-----------|------------|-----------|--------------|-----------|-----------|--------------------------------------|
| MPt, LogP, HD                  | -         | 1.23       | -         | 0.51         | -         | 0.99      | Log P > HD > MPt                     |
| MW, MPt, Log P, HD             | 5.22      | 1.28       | -         | 0.51         | -         | 1.03      | Log P > HD > MPt > MW                |
| MW, MPt, SP, Log P, HD         | 5, 20     | 1.27       | 31.09     | 0.51         | -         | 1.0       | Log P > HD > MPt > MW > SP           |
| MPt, Log P, HA, HD             | -         | 1.14       | -         | 0.85         | 2.51      | 1.11      | Log P > HD > MPt > HA                |
| MW, SP, Log P, HD              | 0.77      | -          | 83.70     | 0.64         | -         | 0.62      | HD > Log P > MW > SP                 |
| MPt, SP, Log P, HD             | -         | 1.22       | 24.47     | 0.51         | -         | 0.98      | Log P > HD > MPt > SP                |
| MW, LogP, 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                   |

TABLE 4  
THE BEST FEATURE SETS WHEN THERE IS A RESTRICTION IN HOW MANY CAN BE USED

| <b>Lowest NMSE model</b> | <b>GP models</b>           | <b>NN models</b>           | <b>GP NMSE ± SD</b> | <b>NN NMSE ± SD</b> |
|--------------------------|----------------------------|----------------------------|---------------------|---------------------|
| 2 features               | MW, HD                     | MPt, HA                    | 0.77<br>± 0.19      | 0.91<br>± 0.13      |
| 3 features               | MPt, Log P, HD             | MPt, SP, HA                | 0.64<br>± 0.13      | 0.91<br>± 0.14      |
| 4 features               | MW, MPt, log P, HD         | MW, MPt, SP, HA            | 0.65<br>± 0.15      | 0.93<br>± 0.18      |
| 5 features               | MW, MPt, SP, Log P, HD     | MW, MPt, SP, HA, HD        | 0.65<br>± 0.14      | 0.96<br>± 0.19      |
| 6 features               | MW, MPt, SP, Log P, HA, HD | MW, MPt, SP, Log P, HA, HD | 0.71<br>± 0.15      | 0.99<br>± 0.20      |

Next we look at the best combination of features if the number of features allowed is restricted. The results are

presented in Table 4. Unsurprisingly it can be seen that the GP does better than the linear model. For the GP only 3 features were needed for best performance. Adding further features did not change the performance much, although performance did drop when the sixth feature (*HA*) was added. Once again *Log P* and *HD* were important features here and this is consistent with the length scale results.

For the linear predictor no improvement was made above just using two features. Interestingly the linear models did not want to use *log P*, presumably because the relation between lipophilicity and permeability is highly non-linear.

## VI. DISCUSSION

Firstly we have confirmed and generalized our earlier result: GPs perform much better than linear predictors on skin permeability prediction. This is true regardless of which features are used as inputs to the model and simply indicates that the relationship between molecular properties and permeability is non linear.

For this problem domain the data set is very small and the number of molecular features that can be used is also small. This allowed us to undertake an unusual study. We were able to look at *every* combination of the six input features to find which worked well. We found that there was no simple ordering of the features, although one of the features, solvability, was consistently the least informative. Different sets of features produced predictions that were equally good, although some features, for example lipophilicity, were present in all the best performing predictors. For others their importance varied with the varying features they were paired with.

The best performing model we could find was a GP with just 3 inputs, although having more features did not, for the most part, damage performance.

We have recently been working on predicting the permeability of skin of non-humans (for example pig skin). We next want to find out if it is possible to improve permeability predictions of human skin with permeability data from other species.

## REFERENCES

- [1] M. T. D. Cronin, J. C. Dearden, R. Gupta, and G. P. Moss, "An investigation of the mechanism of flux across polydimethylsiloxane membranes by use of quantitative structure-permeability relationships.," *Journal of Pharmacy & Pharmacology*, vol. 50, pp. 143-152, 1998.
- [2] I. T. Degim, "New tools and approaches for predicting skin permeability," *Drug Discovery Today*, vol. 11, pp. 517-123, June 2006 2006.
- [3] G. L. Flynn, "Physicochemical determinants of skin absorption," T. R. Gerrity and C. J. Henry, Eds.: Elsevier, New York, 1990, pp. 93-127.
- [4] G. P. Moss, Cronin M.T.D, "Quantitative structure-permeability relationships for percutaneous absorption: re-analysis of steroid data," *International Journal of Pharmaceutics*, vol. 238, pp. 105-109, 2002.
- [5] G. P. Moss, Y. Sun, N. Davey, R. Adams, W. J. Pugh, and M. B. Brown, "The application of Gaussian Processes to the prediction of percutaneous absorption," *Journal of Pharmacy & Pharmacology*, vol. 61, pp. 1147-1153, 2009.
- [6] R. O. Potts and R. H. Guy, "Predicting skin permeability," *Pharmaceutical Research*, vol. 9, pp. 663-669, 1992.
- [7] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*: The MIT Press, 2006.