Plasma Protein Binding Affinity and Its Relationship to Molecular Structure: An In-silico Analysis

M. Paul Gleeson*

Computational, Analytical & Structural Sciences, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, United Kingdom

Received August 14, 2006

In-silico plasma protein binding (PPB) models have been generated on human and rat in-house datasets, and on a human dataset sourced from the literature. From the results reported herein, it is apparent that models built on datasets relevant to the chemotypes under investigation in lead optimization programs will perform considerably better in this role than those generated on diverse compounds sourced from the literature. The in-house human and rat partial least-squares regression (PLS) models have cross-validated q^2 values of 0.53 and 0.42 on the training sets, respectively. On the independent test and validation sets, they display similar predictive ability, with logK prediction errors of ∼0.5 log units. This compares to ∼0.25 log units variability expected for experiment. Given the considerable interspecies PPB differences, the prediction of PPB in one species using measurements in the other is no better than a prediction from an in-silico model generated on that species.

1.0. Introduction

Plasma protein binding (PPB) is an important biological property that can have implications for a number of toxicological, pharmacological, and pharmacokinetic parameters. The fraction of drug unbound in plasma is considered to be the most important for biological effects given it is only the free drug that can elicit a pharmacological response. The free fraction is equally important for toxicological reasons since reduced PPB, which is known to occur in infants and certain disease states,[1,2] could potentially lead to plasma levels above the maximum tolerated dose.

Pharmacokinetic and pharmacodynamic properties heavily depend on PPB. Using the well-stirred model,[3] the in vivo hepatic clearance (Cl_H) can be obtained from intrinsic clearance (Cl_I) and the fraction unbound in plasma (fup) (eq 1), where Q_H refers to the liver blood flow of the species in question. The apparent volume of distribution at steady state (V_{ss}) is also dependent on PPB according to the Gillette equation[4] (eq 2). This is expressed using the volume of the plasma compartment (V_p), the volume of the tissue compartment (V_t), and the fraction unbound in both plasma and tissue (fut).

$$Cl_H = \frac{Q_H \cdot \text{fup} \cdot Cl_I}{Q_H + \text{fup} \cdot Cl_I} \quad (1)$$

$$V_{ss} = V_p + V_t\left(\frac{\text{fup}}{\text{fut}}\right) \quad (2)$$

Highly protein bound drugs have extra developmental hurdles, with the FDA requiring additional studies. This of course is because small changes in the %bound can have a significant impact on free fraction of drug in the body. For example, the 4% difference between 99% bound and 95% bound corresponds to a 5-fold difference in the %free, whereas the 4% difference between 80% bound and 76% bound is only a 1.2-fold difference. Given the obvious importance of PPB in drug discovery, it would be extremely useful to be able to rapidly estimate or rank the extent of binding to plasma proteins early in the life of a discovery program.

A number of groups have investigated the statistical relationship between molecular structure/properties and human serum albumin binding[5,6] and to the more physiological relevant whole plasma.[7−9] The latter matrix contains multiple proteins, including the two most abundant albumin and alpha glycoprotein. A common feature of these analyses was the relatively small, primarily literature based datasets employed, derived from a variety of laboratories, from different assay types (A description of these techniques can be found in the following references: equilibrium dialysis,[10] ultrafiltration,[11] and HPLC[12]). Furthermore, the datasets used consisted of primarily marketed compounds that are not necessarily ideal to develop models to predict the current generation of hits and leads under investigation within the pharmaceuticals industry due to the considerable differences in molecular properties.[13,14]

The binding of large numbers of compounds to plasma have been determined in GSK as part of ongoing lead optimization efforts. These measurements are predominantly performed using an equilibrium dialysis assay, in either plasma and to a lesser extent whole blood, as well as a number of preclinical species, including human, rat, dog, and guinea pig. This gives us a unique opportunity to study the relationship between PPB and molecular structure, as well as any differences that arise between species, using a dataset that will give statistically meaningful results. The study reported here is limited to equilibrium dialysis data in rat and human as these represent the two most important, most frequently assayed species. Data from the paper of Yamakazi et al.[8] were used for the purpose of comparison. Univariate and multivariate statistical methods were considered for the analyses so as to afford a set of simple, generic rules to reduce PPB, as well as a more complex in-silico model to predict the extent of binding or at least to provide a rank ordering of compounds from individual program series.

For a QSAR/QSPR model to be applied with any confidence, it must demonstrate a predictive ability on an independent test set of compounds not used in the building of the model. However, a random or factorially selected test set is not the most effective measure of a model's performance since it is

* To whom correspondence should be addressed. Phone: +44 (0)1438 768682. Fax: +44 (0) 1438 763352. E-mail: paul.x.gleeson@gsk.com.

10.1021/jm060981b CCC: $37.00
Published on Web 12/15/2006

essentially a mirror image of the original training set. A more challenging temporal validation set is used for the models built in GSK, collected for at least 6 months after the model was completed. This is generally a more effective measure of the model performance because new drug discovery programs come on line periodically, and the chemical space of existing programs typically increases as teams try to optimize their leads to obtain more potent, selective, less liable compounds. Finally, a consideration of how the similarity of a query compound compares to those used to build the model (i.e., the distance to model) and its relationship with the prediction error is assessed. This can provide useful information on how well one might expect the model to perform on data that are found far from the training set chemical space.

2.0. Results and Discussion

2.1. Univariate Analysis Results. It has been reported in certain literature sources that protein binding to plasma[7,8] or human serum albumin (HSA)[5,6] is heavily influenced by the logP/logD term. In another the term is said to be unimportant globally,[9] while in yet another, it is stated to be unimportant within individual series.[15] Less controversial is the finding that acids are generally more highly bound to plasma proteins than neutral or basic compounds given the greater concentration of the fatty acid transporting protein, albumin, compared to alpha glycoprotein, which generally prefers bases.[12] To try to clarify these reports, the results from the univariate analyses based on our large, pharmaceutically representative dataset are discussed.

One of the key physicochemical characteristics affecting the extent of PPB, namely, the ionization state, can be appreciated from the ANOVA results on 1,435 compounds measured in rat plasma. Analysis of the mean logK values of the different ionization states (Figure 1a) reveals that acidic compounds have considerably higher mean logK values than either neutral, zwitterionic, or basic compounds. Zwitterionic and neutral compounds are found to have comparable mean logK values, although the small number of zwitterions means the 95% confidence limits in its mean are considerably larger, making it difficult to say whether there is a real difference between the two types. Bases make up ~60% of the dataset, and these have the lowest mean logK of all ionization types. Even though the mean difference in logK between bases and acids is ~1 log unit, from the standard deviation of the former compounds, we can see that in certain circumstances they can have as high an affinity for plasma proteins as acids (Figure 1b).

The results displayed in Figure 1a,b do not take into account whether other factors important in protein binding, such as logP, differ significantly within each ionization group, which could potentially give rise to misleading results. To answer this question, the effect of lipophilicity on top ionization state was assessed (Figure 1c). Increasing lipophilicity leads to an almost constant increase in the mean logK irrespective of the ionization type, indicating this parameter has a significant positive effect on binding. Only for acids in the 3−5 and >5 logP bins does this constant effect appear to break down. However, again the 95% confidences in the means are quite large for acids as a result of the relatively small number of observations. Thus, for a given logP it is apparent that acids are more highly bound to plasma proteins than neutrals, which are in turn more highly bound than zwitterions or bases. From Figure 1c, we can see that zwitterions generally have a lower logK value than neutrals for a given logP.

The relationship between lipophilicity and PPB for the four individual ionization states can be displayed in a more traditional

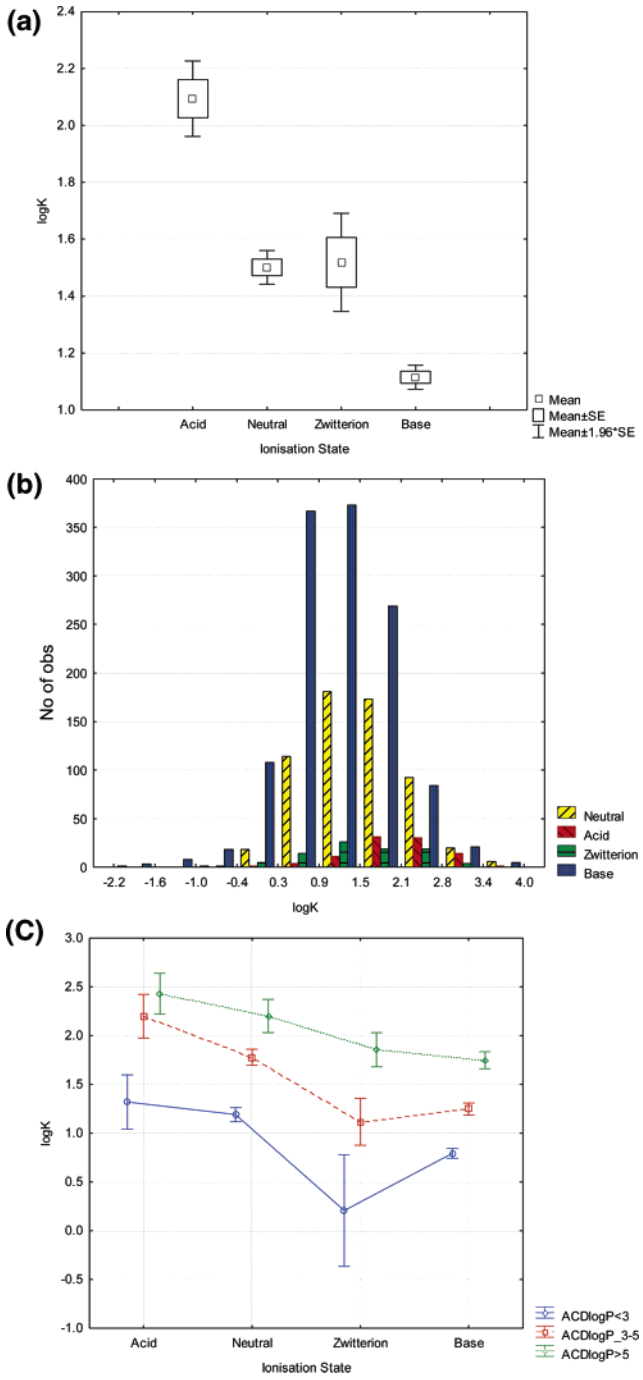

Figure 1. (a−c) Relationship between PPB, ionization state, and logP for all measured compounds in rat. (a) Mean logK values illustrated by ionization state, (b) distribution of logK values for a given ionization state, and (c) mean logK values illustrated by ionization state for a given logP range. The one-way ANOVA results associated with (a) and the two-way ANOVA results associated with (c) are statistically significant above the 95% confidence level.[32]

form using linear regression (Figure 2). Of the two lipophilicity parameters considered, ACD logD at pH 7.4 and ACD logP, the latter was found to be more strongly related to logK. logP describes a highly significant portion of the variability in binding to plasma proteins, ranging from 24% of the total variation for bases (i.e., $r^2 = 0.24$), 27% for neutrals, 38% for acids, and 58% for zwitterions. This compares to just 21% when all ionization states are considered together. The slopes and intercepts of the regression lines given in the caption of Figure 2 show that a log unit increase in logP will lead to a similar increase in logK for acids, bases, and neutral compounds with

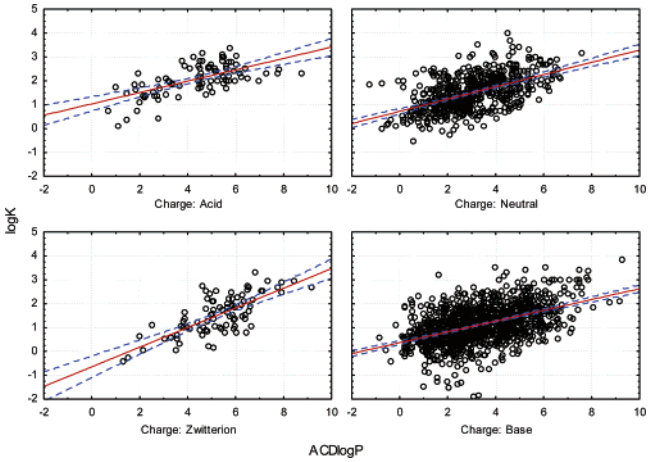

Figure 2. Relationship between rat logK and ACD logP for the four different ionization states: acid ($r^2 = 0.38$, intercept (I) = 1.89, slope (M) = 0.09); neutral ($r^2 = 0.27$, $I = 1.04$, $M = 0.16$); base ($r^2 = 0.24$, $I = 0.54$, $M = 0.26$); zwitterion ($r^2 = 0.53$, $I = 0.37$, $M = 0.34$). For all ionization states combined ($r^2 = 0.24$, $I = 0.23$, and $M = 0.50$). Regression lines are in red and 95% confidence limits are in blue.

slopes of 0.24, 0.23, and 0.26, respectively. The regression line for the zwitterionic compounds has a markedly different slope of 0.41, suggesting that a log unit increase in logP will have a greater effect on the logK. This may be an artifact of the relatively small dataset ($N = 58$), even taking into account the larger 95% confidence intervals for the regression line compared to the other types.

The overall correlation between logP and logK may not seem particularly strong; however, this is simply a reflection of the diversity of the dataset and the fact that the binding process is not controlled by bulk properties alone. For example, multiple molecules can bind simultaneously to the multiple sites available on a single albumin protein which we cannot account for. In fact, lipophilicity as given by ACD logP is found to correlate more strongly with logK than any of the alternative descriptors listed in Table 1, and while the correlations observed in Figure 2 may be relatively weak they are all statistically significant above the 95% confidence level.

To emphasize the importance of lipophilicity, the rat dataset was broken down by individual program series and reanalyzed. The program datasets discussed here typically consist of a single structural template and ionization type (i.e., individual series), which reduces the likelihood of confounding effects due to molecular recognition differences or other factors (Table 2 and Figure 3). In these examples, it is found that ACD logD at pH 7.4 correlates more strongly than ACD logP, possibly as a result of the subtly differing degrees of ionization found within each subset. In Table 2, the line of best fit from the relationship between logK and logD is used to calculate the error in prediction. One can then use the errors along with the r^2 to assess the importance of logD for each program.

A very strong relationship between logK and the calculated logD value is found for programs A and B. These are as good as could be expected given the variability in the experimental assay (~0.25 log units), indicating that lipophilicity alone governs their difference in affinity. One should note that the RMS error is used here as the primary statistic to assess the relationships rather than the r^2, as the latter is heavily dependent on the variance of logK values of the particular set being studied (σY). For example, program F has a larger logK variance than does E (σY of 0.66 compared to 0.42), yet based on the relationship as defined with logD one can predict the logK value

Table 1. Summary of the Descriptors Used in This Study

variable	description
abe	Andrews binding energy
cmr	calculated molecular refractivity
mw	molecular weight
tpsa	polar surface area (Ertl)
aring	no. of aromatic rings
naring	no. of nonaromatic rings
hba	no. of H-bond acceptors
hbd	no. of H-bond donors
neg	no. of negatively ionizable/charged groups
pos	no. of positively ionisable/charged groups
flex	ratio of the number of rotatable bonds to total bonds
alpha	the overall or summation solute hydrogen bond basicity
betaH	the overall or summation solute hydrogen bond acidity
R2	the excess molar refraction
Pi	a combined dipolarity/polarizability descriptor
vx	McGowan's volume
clogp_acd	calculated ACD logP
%acidic_form	% of the molecule in the acidic form at pH 7.4 using ACD pKa
%basic_form	% of the molecule in the basic form at pH 7.4 using ACD pKa
%zwitter_form	% of the molecule in the zwirtterionic form at pH 7.4 using ACD pKa
%neutral_form	% of the molecule in the neutral form at pH 7.4 using ACD pKa
total_HB	sum of H-bond donors and acceptors
total_charge	sum of positive and negative charges
%hba	% of H-bond acceptors/total number of atoms in molecule
%hbd	% of H-bond donors/total number of atoms in molecule
logd_pH20_acd	calculated ACDlogD at pH = 2.0
logd_pH55_acd	calculated ACDlogD at pH = 5.5
logd_pH65_acd	calculated ACDlogD at pH = 6.5
logd_pH74_acd	calculated ACDlogD at pH = 7.4
logd_pH110_acd	calculated ACDlogD at pH = 11.0

Table 2. Relationship between PPB and ACD LogD at pH 7.4 for the Best Correlating Programs[a]

program	r^2	RMSE[b]	σY	slope	intercept	N
A	0.84	0.18	0.47	0.55	−0.42	12
B	0.84	0.24	0.61	0.63	−0.59	30
C	0.62	0.31	0.52	0.40	0.07	18
D	0.67	0.32	0.58	0.52	−0.84	12
E	0.41	0.32	0.42	0.22	0.88	78
F	0.68	0.37	0.66	0.43	−0.08	20
G	0.60	0.49	0.8	0.45	−0.52	16
H	0.45	0.50	0.68	0.37	0.54	38

[a] Programs are sorted by decreasing prediction error (RMSE) rather than r^2. [b] Fitted RMSE as calculated from the regression equation.

of F less accurately than E (0.37 log units vs 0.32) even though it has the larger r^2 (0.68 vs 0.41). For programs C to H, the rank ordering of compounds using logD is still reasonable; however, the amount of explainable variance in the observed value decreases. This is most likely due to different confounding factors that are not present in programs A and B.

The confounding factors would also help to explain why the relationship between logD and protein affinity is markedly different for different programs, as can be seen from the different slopes and intercepts (Table 2). For example, the regression results show the logK value of program A (Figure 2a) reduces by 0.55 log units with a log unit drop in logD, while for program E (Figure 2b) it is only approximately 0.2 log units. Furthermore, for a compound with a logP of 0, one finds the predicted logK value of the program E is 0.88 compared to −0.84 for program D.

Table 3. Model Statistics for the GSK Human, GSK Rat, and Literature Human Models[a]

	line of unity statistics			line of best fit statistics			other statistics		
(a) human GSK model	r_0^2 (q^2)	RMSE	ME	r^2	slope	intercept	distance	σY	N
training data	0.56 (0.54)	0.55	0.00	0.56	1.00	0.00	1.00	0.83	686
test data	0.48	0.54	−0.02	0.58	0.95	0.11	1.00	0.83	211
validation data	0.50	0.57	0.07	0.51	0.97	−0.02	0.93	0.81	385
rat data[b]	0.17	0.64	0.23	0.44	0.92	−0.12	1.01	0.78	956
literature data	0.03	1.05	0.6	0.34	0.81	−0.41	1.89	1.07	324
(b) human literature model	r_0^2 (q^2)	RMSE	ME	r^2	slope	intercept	distance	σY	N
training data	0.45 (0.40)	0.83	0.00	0.45	1.00	0.00	1.00	0.92	243
test data	0.46	0.67	0.17	0.51	0.85	0.23	1.02	1.11	82
GSK data	0.29	0.70	0.27	0.42	0.81	0.55	1.75	0.83	1282
(c) rat GSK model	r_0^2 (q^2)	RMSE	ME	r^2	slope	intercept	distance	σY	N
training data	0.44 (0.43)	0.62	0.00	0.44	1.00	0.00	1.00	0.84	1081
test data	0.47	0.58	−0.05	0.48	1.01	0.04	1.02	0.80	347
test data (blood)	0.38	0.57	−0.09	0.40	0.92	0.17	0.87	0.73	614
validation data	0.51	0.63	−0.11	0.56	1.37	−0.58	1.22	0.89	624

[a] Line of unity statistics: r_0^2 is the correlation coefficient to the line of unity, q^2 is the cross validated r_0^2 from a leave many out procedure, RMSE is the root-mean-square error in prediction and ME is the mean error in prediction. Line of best fit statistics: r^2 is Pearson's correlation coefficient squared, and the slope and intercept define the line of best fit equation. Other statistics: distance is the average distance of the dataset to the origin of the scores plot (absolute sum of the t values for all components), σY is the standard deviation of the observations LogK values and N is the total number of observations. [b] Human model predicting the non-overlapping subset of rat data also shown.

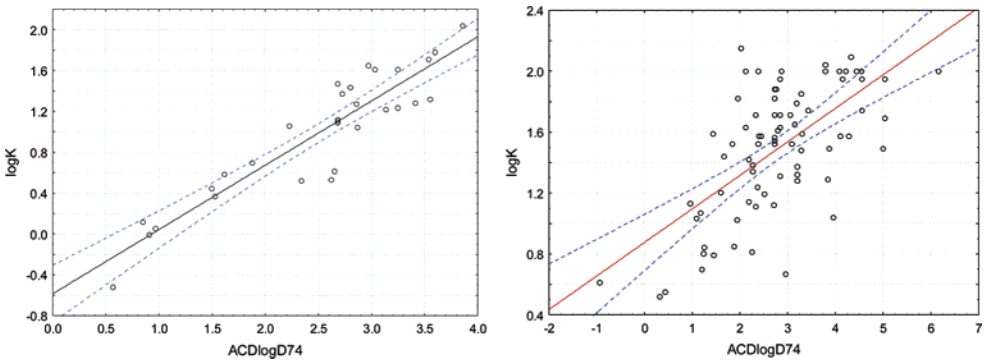

Figure 3. Relationship between PPB and logD broken down by program (A left & E right). For a number of programs, one can find a strong correlation with logD alone. The correlation for the program A in the top left is as good as the experimental reproducibility of the assay meaning logD alone governs the difference in binding.

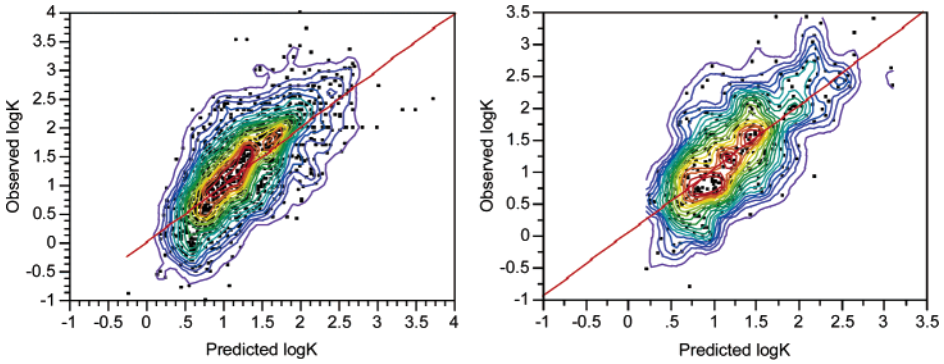

Figure 4. Plot of observed versus predicted human training (left) and test set (right) results. Red contours represent area of maximum density, while blue represents the lowest.

2.2. Partial Least-Squares (PLS) Regression Results. To profile the relationship between the extent of binding and molecular structure more comprehensively, one must construct more sophisticated multivariate statistical models that can cope with multiple input parameters such as logP, ionization state/extent, size, polar surface area, hydrogen bond donor and acceptor strength, etc. Since many of the descriptors are correlated, partial least-squares regression is employed since it can cope with this and also provide greater interpretability than commonly used alternatives such as neural networks or ensemble classification/regression trees.

2.2.1. Human Model Details. The statistical model built on 686 human PPB measurements has a moderate r_0^2 of 0.56, and an equivalent r^2 since it is a fitted relationship with a slope of 1 and an intercept of 0 (Table 3 and Figure 4a). The corresponding cross-validated q^2 is approximately the same as the r_0^2 at 0.54, suggesting that the model is not overfitted. The prediction error as given by the root-mean-square error (RMSE) is 0.56 log units compared to experimental variability determined at ~0.25 log units. The latter value is based on an analysis of >100 repeat measurements in rat. Given that the standard deviation in logK for the training set is 0.83, the maximum

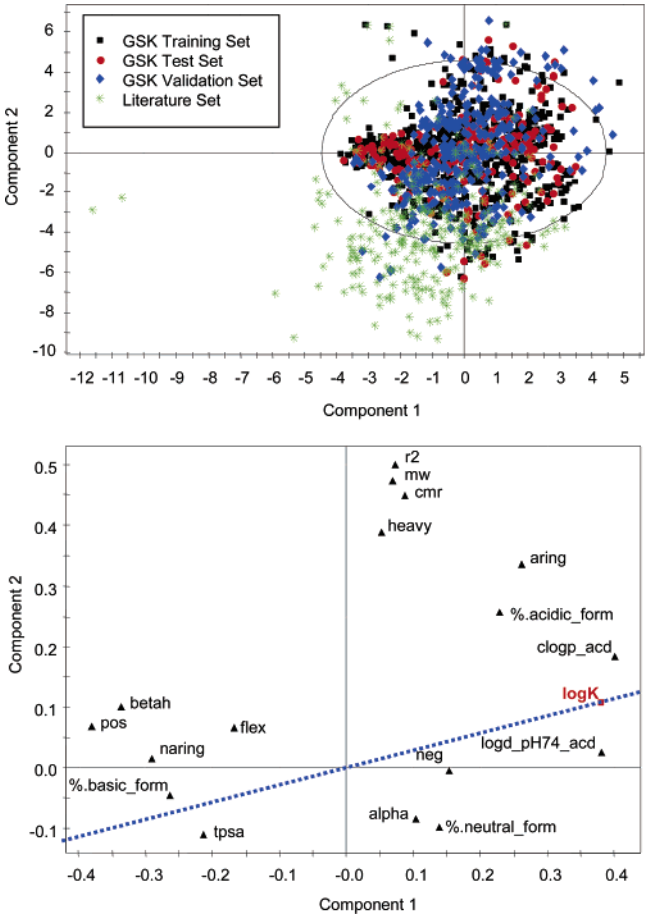

Figure 5. Scores plots showing component 1 (t1) versus component 2 (t2) (top) and the corresponding loadings plot (bottom). The distance to the training set in PLS model space are 0.89 (self-similarity), 0.90, 1.06, 1.68 for the training, test, validation, and literature set, respectively. Two extreme outliers are found on component 1 having very low logDs due to their five basic centers (gentamicin & tobramycin). Descriptors at the extreme ends of the blue line in the loading plot have the greatest impact on the PLS model and thus the distance. Only the first two components are displayed as they describe ~80% of the total model variance.

possible r_0^2 one could expect is 0.91, meaning the in-silico model describes ~60% of the explainable variance in the training set.

The model performance on the randomly chosen test set of 162 compounds is a better guide to the predictivity of the model than the training set itself since it has not had an influence on the model-building process. The results show the model is of moderate strength, and is not overfitted, since the overall correlation and prediction errors are essentially equivalent in the two sets (Table 3 and Figure 4b). Since the test set is in essence a mirror image of the training set due to the random selection employed here, the comparable average distance to model is understandable. The distance to model is calculated to be approximately 1 on average, indicating the test set lies within the model space. While this distance represents a 1D representation of the overlap between the datasets, we can see the same effect in 2D by projecting the compounds into the PLS component space (Figure 5).

Such a good performance on the test set might be expected given the observed overlap. However, a sizable discrepancy between the r_0^2 and r^2 statistics in the test set is observed reflecting the fact that the slope and intercept of the line of best fit deviate from 1 and 0, respectively. Even so, the prediction error is comparable to the training set with only a small net overprediction of the data (ME = −0.02).

The 385 compound temporal validation set available to assess the model is found to have a distance to model of 0.94. This is less than that of the training set indicating in this case that the newly synthesized data have not strayed significantly from the model space as defined and so might be expected to be well predicted by the model. The r_0^2 for the validation set is found to be ~0.5, the RMSE has increased only marginally (0.03 log units) compared to the training set, and the dataset is on average overpredicted by 0.07 log units. While not the most rigorous test of the model, the validation set shows the model is quite robust to newly synthesized data.

Measurements extracted from the literature were then assessed. The results in Table 3 show that the model performs considerably differently compared to the GSK human model. The RMSE is dramatically larger at ~1 log unit compared to ~0.55 log units for the in-house data. This arises as a result of a 0.60 log unit underprediction of the literature data. This means the correlation to the line of unity is quite poor, although the model does have some ranking ability as given by the r^2 of 0.34. From an analysis of the PLS scores plot (Figure 5), one can see that the GSK compounds occupy a significantly different area of property space compared to the literature set which might explain the results. Marketed compounds, which predominate the literature dataset, have considerably lower mean molecular weights and logP's than the compounds in the GSK dataset. Furthermore, acids dominate the literature set, an artifact of the greater need to report their PPB, even though acids make up a small portion of drugs.

For 693 compounds with experimental PPB measurements in both human and rat, we find a moderate correlation between the values. However, the mean experimental logK is 1.44 log units in human, 0.11 log units larger than that found in rat. Even taking into account experimental variability of ~0.25 log units, the probability of observing such a large difference in the mean logK values of the two species is statistically highly unlikely given the number of observations. Applying a paired t-test to the data shows that this is a significant difference above the 95% confidence level. Furthermore, this finding is in agreement with the analysis of a similarly sized AstraZeneca dataset by Gleeson et al., suggesting that this is a real difference between the species.[13]

We can therefore use the non-overlapping set of rat measurements to assess the predictive ability of the human model. One might expect the human model to rank the rat data reasonably well but not necessarily predict the absolute logK accurately. Furthermore, since the distance to model for this set will be comparable to the well-predicted human test and validation set, one might also expect the model to perform well on the data, albeit reflecting the experimental differences between species. The results in Table 3 show that rat data are poorly predicted by the human model in absolute terms, with an r_0^2 of 0.17. The line of best fit statistics are better (r^2 = 0.44), a result of the data being underpredicted by an average of 0.23 log units. This difference is in the correct direction but larger in magnitude than the experimental differences between species.

Distance to Model versus the Error in Prediction. The effect of distance to model space for the available independent datasets was assessed in more detail. The test set data were first binned into quartiles (four equally sized bins), and the relationship between the distance bin and the error in prediction was assessed. (Figure 6). The RMSE is found to increase by 0.07 log units going from the closest quartile of the test set to the

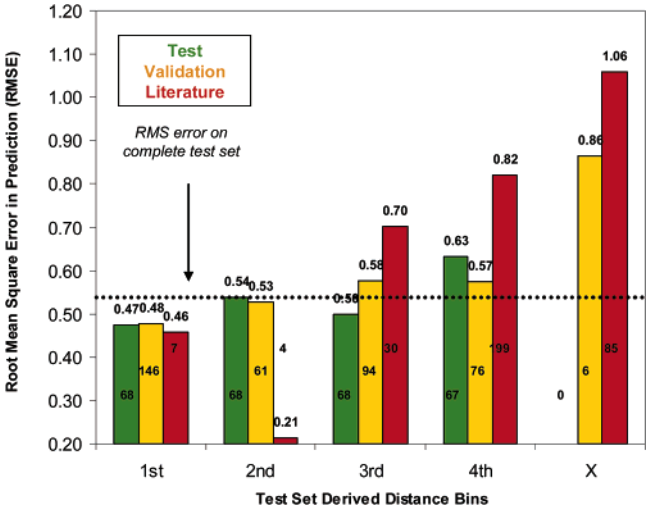

Figure 6. Plot of the RMSE of the test, validation, and literature set using the quartiles associated with the test set. Those compounds that lie outside the maximum value of the test set are found in the distance bin (X).

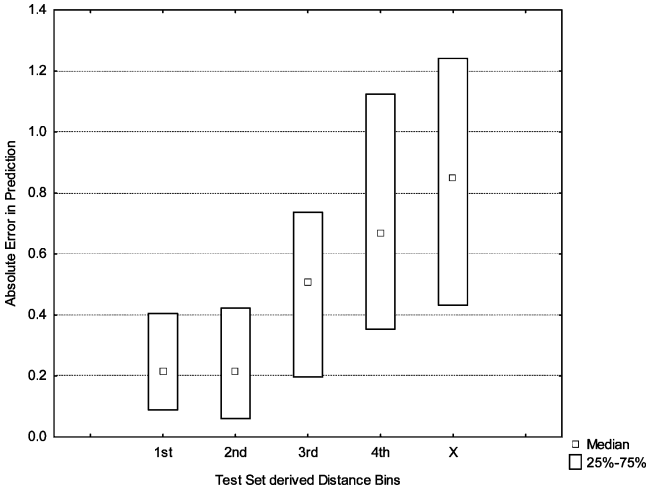

Figure 7. Plot of the absolute error in prediction vs the distance to model for the literature set binned according to the test set quartile distances. The Krustal-Wallis ANOVA by ranks test shows differences in the medians of the literature and validation set to be significant above the 95% confidence levels. A nonparametric test must be used to assess the statistical significance as the data are not perfectly normally distributed. The small square denotes the median, and the rectangular box denotes the range in which 25−75% of the errors lie.

second closest. The RMSE actually decreases by 0.04 log units in the third, but in the final it increases again by 0.13 log units. Thus, apart from the third bin, the RMSE increases with increasing distance. This effect appears more pronounced in the validation set. There is a gradual increase in RMSE with each distance bin in the validation set, but this begins to plateau once the third quartile has been reached. The six observations in the dataset that lie outside the maximum distance of the original test set are the most poorly predicted of all.

In the literature dataset, we only find 11 observations within the first-second distance quartiles, and these are much better predicted than those that lie further away. The majority of the data lies beyond the third quartile and is predicted with considerably less accuracy.

To ensure the finding that the RMSE gradually increases with the distance quartiles, the statistical significance of the relationship was assessed. Using the nonparametric Krustal-Wallis test, one finds the relationship between the number of compounds found above the median error increases with the increasing distance bin for the validation and literature sets, and these are significant differences above the commonly used 95% confidence level (Figure 7). Those for the test set are only significant at ∼50%. Even considering the relatively weak statistics found in the latter dataset, these results show that a knowledge of the domain of applicability can be used to provide information about the reliability of a prediction. The ability to define the reliability of a given prediction (i.e., below the RMS model error, equivalent to, or considerably greater than) means predictions can be used with more confidence.

2.2.2. Human Literature Model Details. The reliability of QSAR/QSPR models reported in the literature are often questioned[13,14] since they are typically generated on compounds with markedly different physicochemical properties and were determined from assays with often distinctly different protocols. The results from our literature-based models predicting GSK data are now reported to add further to the debate. The literature model generated here employed the same theoretical protocols as the GSK-based models, so any differences are simply a reflection of the datasets rather than the model-building methodology used.

The results of the model in fit and prediction are shown in Table 3. The training set r_0^2 obtained from the 243 observations is 0.45 compared to 0.56 for the GSK model. The latter was built on 2.5 times the number of datapoints. The literature model's q^2 also shows greater variability from the cross-validation procedure than the GSK model and the RMSE is ∼0.3 log units higher. These collectively indicate the model is less predictive.

The independent test set of 82 compounds lies within the domain of applicability of the model and is moderately well ranked ($r^2 = 0.51$), but the logK value is underpredicted by ∼0.17 log units. This means the correlation to the line of unity is just 0.46. Surprisingly, the test set shows a lower RMSE compared to the training set (0.67 vs 0.83), but this is still larger than the GSK model. This could be an artifact of the small dataset and/or the greater variability in the literature PPB data.

The ability of the literature model to predict the complete human GSK dataset was assessed next. Analysis of the domain of applicability indicates this data lies far from the training set data, suggesting that the data may be poorly predicted. The GSK dataset is found to be moderately well ranked by the literature model ($r^2 = 0.42$), but the net underprediction by 0.27 log units means the r_0^2 is just 0.29. The RMSE is ∼0.13 log units higher than that reported for the GSK model. This seems to confirm that literature-based models should not be applied blindly without a consideration of distance to model.[13] Even if observations lie within the domain of applicability of such a model, the rank order of a set of compounds rather than the absolute prediction should be used as the prediction errors are likely to be considerably larger.

The distance to model can only be used in a crude fashion for the literature model, with those compounds found outside the literature model space being less likely to be well modeled than those within. It was not possible to demonstrate a more precise, statistically significant relationship between the distance and the error for compounds found within the training set space as was demonstrated for the human GSK model.

2.2.3. Rat Model Details. The rat dataset is the largest available to us since this is the most important preclinical species used in the pharmaceuticals industry. Approximately 4 times the number of compounds are available to generate a model

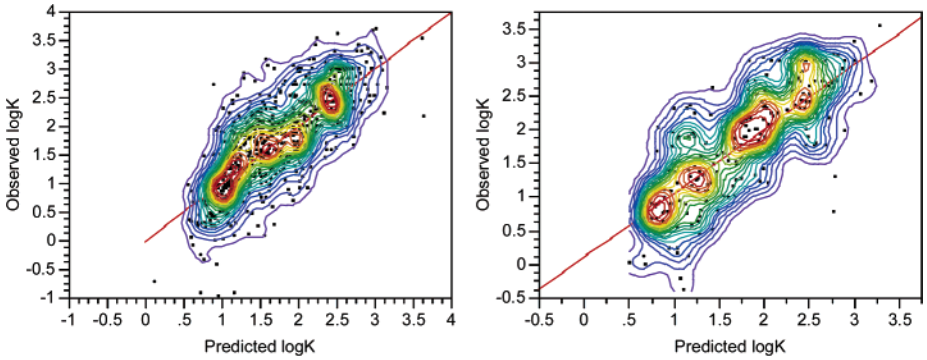

Figure 8. Plot of observed versus predicted rat training (left) and test set (right) results. Red contours represent area of maximum density while blue represents the lowest.

compared to the literature based dataset, which may have implications for the domain of applicability.

The training and test set statistics for the rat model are reported in Table 3 with a graphical illustration of their performance in Figure 8. The in-silico model describes 44% of the variance of the training set and has a cross-validated q^2 of 0.43. This indicates the model is robust even if only moderately predictive. The RMSE of the model is 0.62 log units, larger than the human GSK model but less than the literature human model. The independent test set is predicted to an equivalent degree of accuracy, having an r_0^2 of 0.47 and an RMSE of 0.58. The mean error in prediction is small at −0.05 log units.

The 614 blood-based protein binding measurements are overpredicted to a greater extent than the plasma test set (ME = −0.09), potentially exacerbated by experimental differences. The dataset is ranked with moderate accuracy ($r^2 = 0.40$), and the RMSE is essentially equivalent to the test set. The average distance to model of the blood data is just 0.87, indicating that it lies well with the training set.

The temporal validation set displays a markedly different distance to model distribution, lying further from the training set compared to the randomly partitioned test set (1.02 vs 1.22). The dataset is predicted well ($r_0^2 = 0.51$), although this larger r_0^2 is somewhat a reflection of the larger variance in the dataset. The prediction error is comparable to that of the training set (0.63 log units), but the data are also underpredicted by −0.11 log units on average. Collectively, the results on the independent test and validation sets indicate that rat data are predicted with a lower accuracy than are human data. Furthermore, no human data were available to assess this model as all the compounds measured in human have also been measured in rat.

No statistically significant relationship could be demonstrated between the distance and the error. This may be a result of the larger dataset making the model more stable or the fact that the datasets used to assess the effect are not diverse enough to stress the model. This is not necessarily a problem, potentially meaning the model is more generalizable.

2.2.4. Model Performance by Individual Program Series. In-silico models built on diverse, global datasets can have considerable performance differences on the individual programs or series of compounds found within the global model. In this section, a consideration of how successfully programs from the rat test and validation set are predicted by the global model. This analysis was limited to the rat model due to the larger number of observations. The best predicted programs where the number of observations was greater than 10 were assessed so as to ensure statistically meaningful results could be obtained (Table 4 and Figure 9). All correlations discussed are statistically significant.

Table 4. Rat Model Performance for the Most Populated Program Series in the Test Set

program	r^2	r_0^2	RMSE	ME	σY	N
1	0.64	0.62	0.28	0.04	0.47	13
2	0.61	0.46	0.35	0.15	0.49	21
3	0.24	0.16	0.42	0.08	0.47	23
4	0.68	0.62	0.43	−0.01	0.71	23
5	0.35	0.31	0.45	−0.09	0.56	15
6	0.29	0.27	0.49	−0.05	0.58	37
7	0.71	0.52	0.51	−0.20	0.75	28
8	0.67	0.49	0.56	0.10	0.82	12
9	0.82	0.52	0.56	−0.11	0.84	11

The nine programs in Table 4 display lower RMSEs than those found for the global test/validation sets. This also means that a similar number of programs are correspondingly more poorly predicted than the global model RMSE. In terms of the rank ordering ability of the models, this is heavily dependent on the variance in the individual programs being assessed, which can range from between 0.47 log units to 0.84 log units (Figure 10). Programs 1, 2, and 4 for example are well predicted by the model, displaying high r^2 values and low RMSEs, indicating the rank ordering of the compounds are strong and the logK prediction is accurate. However, program 2 is underpredicted by 0.15 log units resulting in a weaker correlation to the line of unity. Programs 3, 5, and 6 in contrast are the most poorly ranked programs ($r^2 \leq 0.35$), but they still have some of the lowest RMSEs of the nine programs, and lower RMSEs than the global model overall. In these cases, the variability in the experimental logK values are low. Programs 7−9 are also well ranked by the model as can be seen from Figure 10 and from a consideration of the line of best fit statistics ($r^2s > 0.67$), even though the RMSE approaches that of the global test set. This means the large r^2 is simply a result of the large variance in the data itself and it is not that the logK values are accurately predicted. In fact, these three programs generally have the largest mean errors of all the program series in Table 4.

The program predictions of global, multivariate in-silico models do not generally approach the experimental error of the particular assay, in contrast with the QSPRs shown earlier for logD alone (Table 2). In the latter cases, a single binding mode is likely so that the differences in a single parameter alone, such as lipophilicity, reflect the affinity changes. The deficiency of multivariate QSPR models is that they can cloud what may be a relatively simple relationship for the series as they take into account features in the molecules even if they are not important for the binding mode in question. This is because a multivariate model assigns a single coefficient to each model descriptor, yet as can be seen from Table 2 each program studied has a distinctly different dependency on logD. An individual pro-

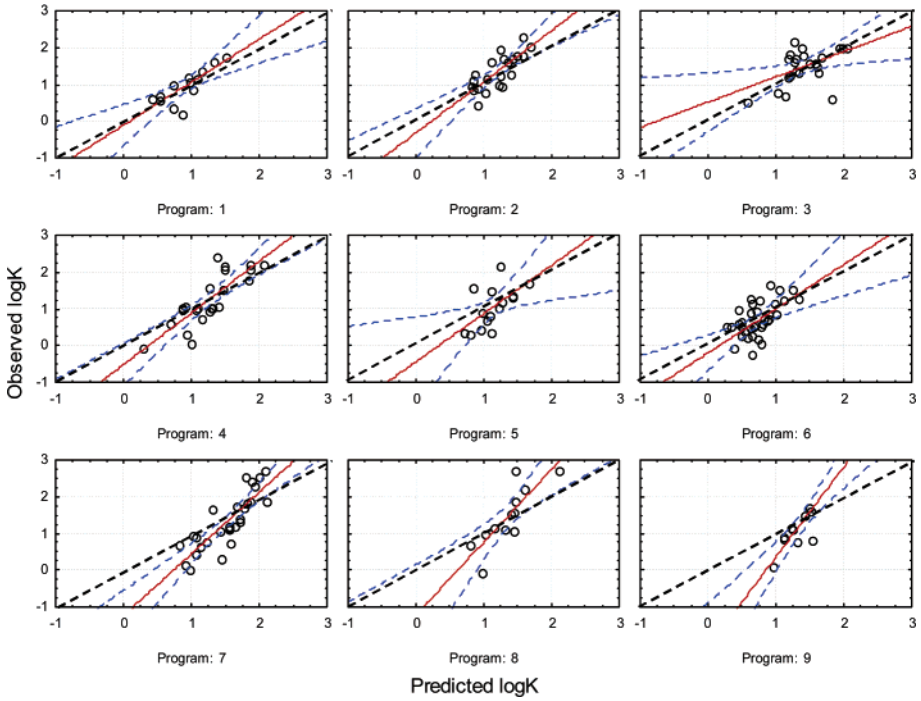

Figure 9. Global rat model prediction by program. The black dashed line indicates the line of unity, the solid red line indicates the regression line, and the dashed blue lines are the 95% confidences in the regression line.

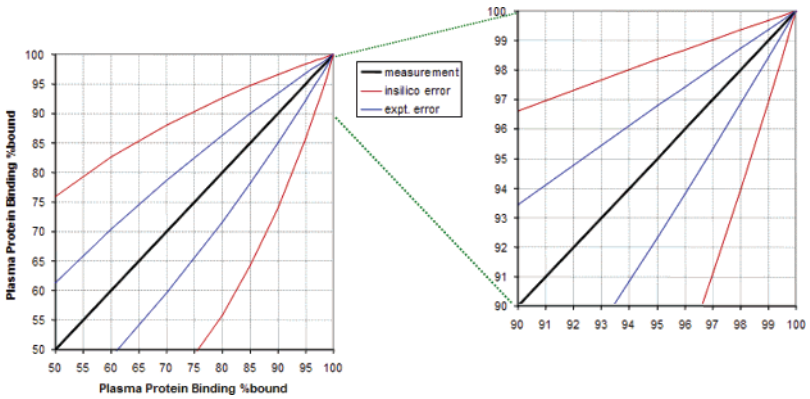

Figure 10. In-silico and experimental errors associated with a given %bound prediction at the 68% confidence level (based on 1 standard deviation of the error - RMSE). The graph to the left describes the 50−100% bound range, and the one to the right shows the more important 90−100% bound range.

gram's dependency on acidity, basicity, or size might also be expected to differ in a similar way making the global model useful but not the optimal solution. However, in the absence of any experimental data the most accurate in-silico model possible is crucial.

Programs within GSK demonstrating a need for greater understanding of PPB may get dedicated computational chemistry support to allow the generation of a local multivariate model, initially using the globally identified descriptors. By essentially reparameterising the global models for individual programs one can often more effectively describe the program data. This however limits the domain of applicability of the model to a particular program only.

2.3. Model Limitations: Experimental Determination vs In-Silico Prediction. It is important one takes into account the model error before any prediction should be used. If one uses the human or rat PPB model in continuous prediction one expects that ∼70% of the time compounds will be predicted within the limits given by the red line in Figure 10. These were calculated by back extrapolating from the logK value and the known model errors. The corresponding blue lines represent the experimental error.

Figure 10 clearly illustrates the relative performances of the two methodologies, the experimental value being considerably tighter than the in-silico prediction. If one wished to know whether a compound had a %bound ≥99% from the in-silico models, a predicted %bound of 95% for example would mean that it is highly likely that the actual value is <99% given the RMSE of the model. This is because one can be ∼70% sure the experimental value would lie between 86.0 and 98.4% based on the model errors. This compares to an ∼70% confidence range between 92.3 and 96.8% bound for the experimental assay. Thus, while this model may not be predictive enough to differentiate the subtle changes required in lead optimization programs, it could be used to rank virtual possibilities.

As a further illustration of the utility of the in-silico models, the performance of the human in-silico model in predicting experimental human binding data was compared to a prediction using experimental values determined in rat alone. Using the absolute rat logK value as a prediction will mean the human

Table 5. Model Statistics for Rat Data Predicting Human (a) Using the Absolute Prediction Untransformed and (b) Using the Line of Best Fit to Transform the Rat LogK Values[a]

expt rat predicting expt human	r_0^2	RMSE	ME	slope	intercept	N
(a) absolute	0.48	0.60	0.22	0.76	0.58	584
(b) regression line	0.62	0.52	0.00	1.00	0.00	584

[a] Data sorted based on increasing prediction error.

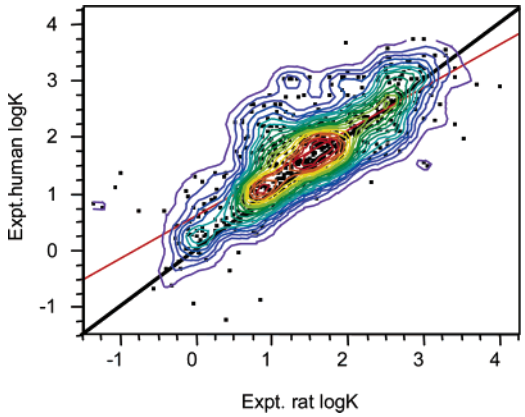

Figure 11. Plot of experimental human logK versus rat. The black line represents the line of unity, and the red line represents the line of best fit. Red contours represent area of maximum density, while blue represents the lowest. The majority of the data lie above the line of unity as a result of the generally greater binding in humans compared to rat.

data are underpredicted by ~0.2 log units with an RMSE of 0.60 log units (Table 5). The rank ordering of the human data is moderate with an r^2 of 0.48. However, it is apparent from Figure 11 that the slope and intercept of the line of best fit deviate significantly from 1 and 0, respectively. The regression equation was subsequently used to correct the experimental rat data to give an optimal human prediction. This best case scenario prediction now has a slope and intercept of 1 and 0, and the RMSE has decreased to 0.52 log units. This however is still considerably larger than the experimental reproducibility of the assay (0.25 log units). Furthermore, the RMSE of the in-silico human model is ~0.55 log units, suggesting using experimental rat data to predict human is no better than the in-silico model built on that species.

2.4. Model Descriptors. We find that three independent in-silico models arrive at similar conclusions regarding the physicochemical factors that control PPB even though they are derived using different training sets and species. The descriptor coefficients of the PLS models, computed as the sum of the influences on each of the fitted components, are displayed in Figure 12. The key descriptor types identified by the model are lipophilicity, size, and charge type/extent of ionization. Size/lipophilicity descriptors are grouped together as they are not totally independent from each other on the PLS loadings plots which can be seen in Figure 5 for the human model built on GSK data. For each of the three models generated here increasing the lipophilicity or conversely decreasing the polarity of compounds generally leads to increasing PPB. Addition of an acidic group or decreasing the acidic pK_a generally leads to increasing binding, while addition of a basic group or an increasingly basic pK_a generally leads to decreasing binding. These results are in agreement with the earlier uni-variate analyses.

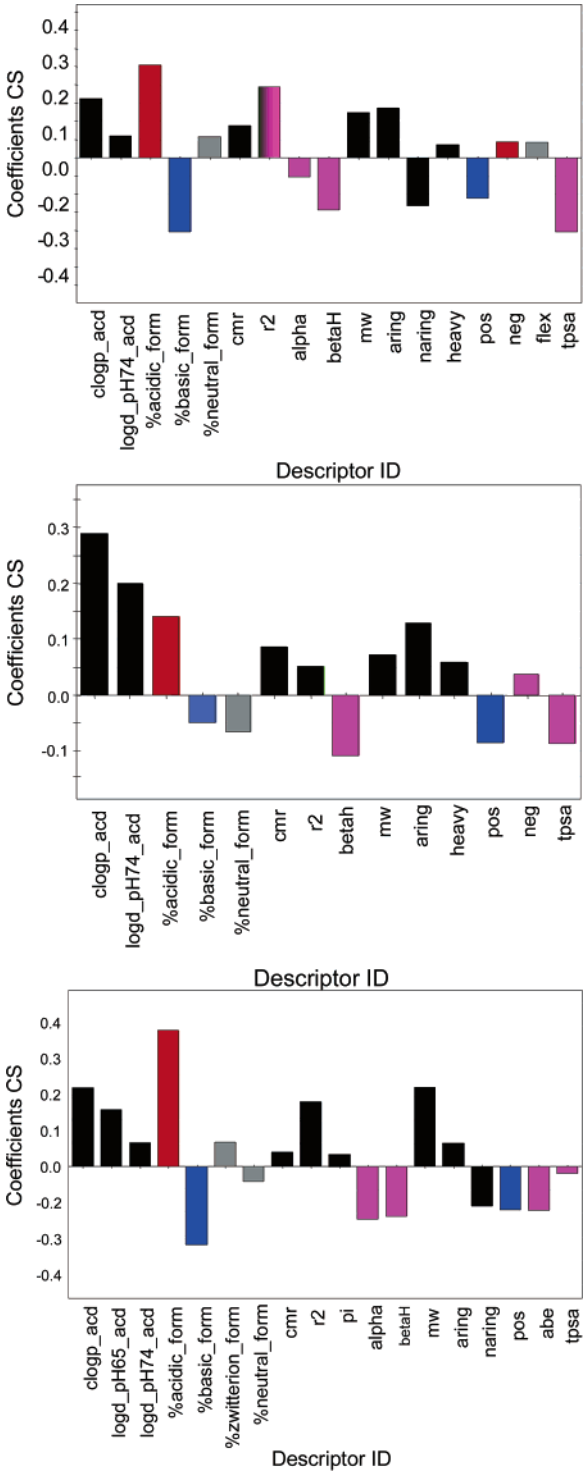

Figure 12. Human (top), literature human (center) and rat PLS model coefficients plot (centered & scaled). Descriptors are colored according to their generic type. Red (acid), Blue (base), size/lipophilicity (black), polar (pink), and others (gray). The models fitted, 4, 2, and 4 components, respectively.

3.0. Conclusions

The results presented herein clearly illustrate the physicochemical determinants of PPB. PPB increases with both increasing lipophilicity and increasing acidity/number of acidic groups. Addition of a basic center or increasing the basic pKa of a molecule will decrease the extent of binding.

In-silico models have been generated on two species, rat and human, and will allow us to effectively rank virtual compounds.

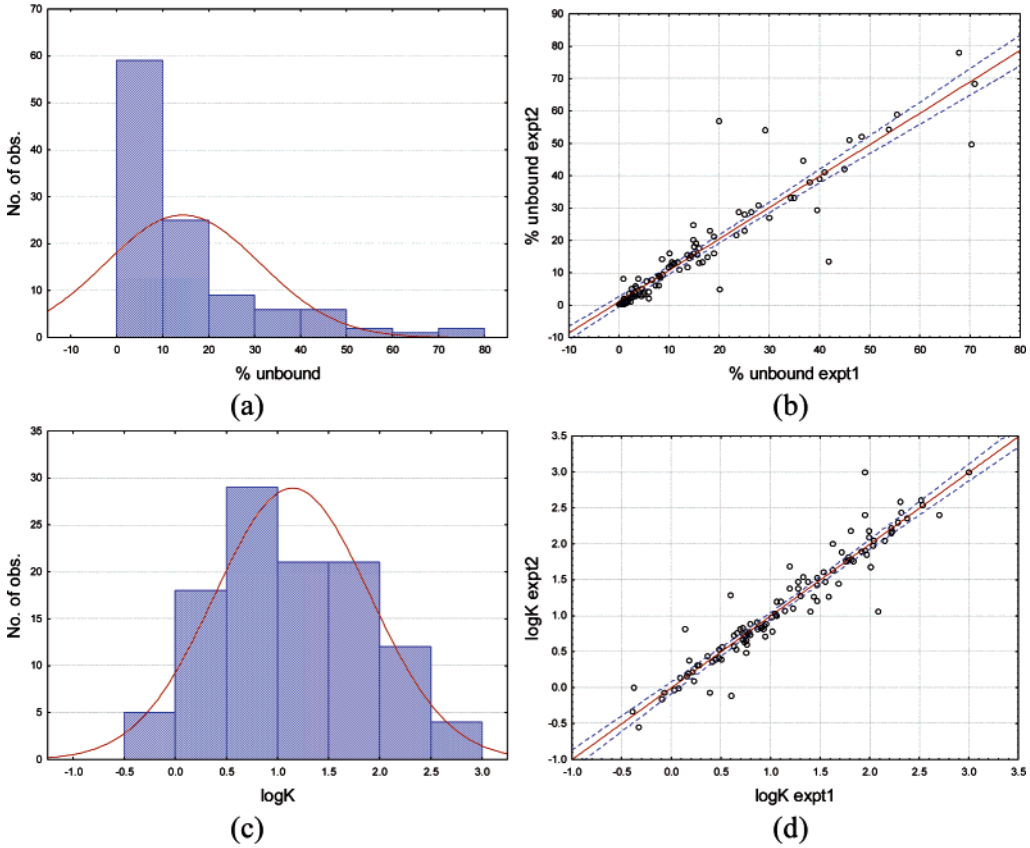

Figure 13. (a−d) Plot of the distribution of untransformed rat PPB data (a) and transformed response ranges (c) and plots of replicate compound measurements on different dates using the untransformed (b) and transformed response values (d). Analysis of the residuals from (b) shows the residuals increase with increasing %free in contrast to (d). Rat equilibrium dialysis, plasma data, $N = 110$.

Furthermore, these models can predict the extent of PPB with the same accuracy as an extrapolation from one species to another. Should a program series be poorly predicted by the global model it is possible to generate more accurate predictions by either re-parametrizing the global model or using correlations derived from individual parameters such as logP.

It is important to consider the extent of overlap in chemical space between the compounds used to train the model and those being predicted. This has implications for models built on limited literature datasets which are then to be applied to different chemotypes types or vice versa.

4.0. Experimental Procedures

A dataset consisting of ∼1500 compounds with equilibrium dialysis (ED), plasma protein binding (PPB) measurements in rat, and ∼900 compounds with measurements in human were randomly split into training (∼75%) and test (∼25%) sets. To test the rat model, 614 ED blood protein binding measurements were available. Generally, blood based data show a good correlation with that derived in plasma, although cases do exist where quite dramatic differences can be found. For test purposes, a human literature dataset of ∼400 compounds was taken from Yamazaki et al.[8] Following the completion of the model building exercise, 6 months of further data were compiled to provide a final, more rigorous temporal validation set.

To assess the predictive ability of our in-house in-silico models compared to one built on a diverse external dataset, a third model was built on the literature dataset extracted.[8] Of the ∼400 compounds reported, 327 remained after structures were obtained and descriptors were calculated. This dataset was partitioned into a training and test set in the same way as the in-house datasets. The predictive ability of this model was subsequently assessed using all of the available in-house human data.

It is important to emphasize the need to normalize the PPB data as commonly reported, before any statistical analysis of the data is performed. The PPB fraction unbound value (fup) or % data required normalization for two reasons. First, analysis of the replicate errors from compounds with multiple measurements shows that the experimental error is not linear across the range, and second, the distribution of the observations is heavily skewed on the fup or % scale (Figure 13a,b). This makes the construction of linear free energy relationships with regression based methods unsuitable. Transformation of the fup into a pseudo equilibrium constant (eq 3), referred to as logK henceforth, leads to a response value with an even distribution of the experimental error across the transformed response range. The observations are now also normally distributed across the response range making regression based methods more reliable (Figure 13c,d).

$$\mathrm{logK} = \log\left(\frac{1 - \mathrm{fup}}{\mathrm{fup}}\right) \tag{3}$$

The %bound, rather than the more appropriate logK value, has been used in a number of statistical models.[9,15] The performance of the model by Yamakazi et al.,[8] while computed as a logK, was reported with statistics based on the %bound, giving artificially inflated r^2 values.

To relate structure to our response values in a way that would be more intuitive to a typical medicinal chemist, the analyses here are limited to a small but comprehensive set of descriptors that describe key bulk property characteristics of molecules that are known to influence almost all DMPK properties. These include lipophilicity at a number of pHs (ACD logD/P[16]), the extent of ionization at pH7.4[17] (ACD pK_a), ionization state indicators, polar surface area,[18] hydrogen bond donor/acceptor indicators and estimations of their strength (Abraham et al.[19]), Andrews binding energy,[20] calculated molar refractivity,[21] etc. (Table 1). Compounds with

missing descriptor values were excluded from the analysis as were compounds that were extreme outliers in any of the descriptors.

4.1. Univariate Data Analysis. An initial assessment was made of the larger set of 1435 GSK compounds measured in rat plasma to discover how the extent of PPB was affected by the two key molecular characteristics often quoted in the literature, namely, lipophilicity and the type/extent of ionization. Analysis of variance (ANOVA) was used to test whether the mean logK values of the four different ionization types[22] (acid, base, neutral, and zwitterions) and/or three equally populated logP bins (<3, 3−5, >5), had significantly different logK values. Consideration of relatively simple, but interpretable, statistical methodologies such as ANOVA is often overlooked in the literature even though these offer clear advantages in terms of interpretability and the opportunity to rationalize simple rules. The relationship between the extent of binding and logP for each of the individual ionization classes was also assessed using linear regression. The extent of the correlation between logK and logP was then assessed by individual program series. All analyses were performed in Statistica 7.0.[23]

4.2. Multivariate Data Analysis. To provide GSK scientists with a means to estimate the extent of protein binding, PLS regression models were generated using the two human and the single rat dataset and ∼40 physicochemical descriptors (Table 1). PLS[24−26] is a least-squares regression technique that (1) assumes a linear relationship between the dependent and independent variables and (2) projects correlated variables in multidimensional descriptor space to lower dimensions. PLS models were initially built in GOLPE[27] fitting four components optimally based on the ratio of the r^2/q^2. Variable reduction was achieved using one round of D-optimal design (30% of redundant variables removed) followed by fractional factorial selection. The models were manually refined further within SIMCA P10[28] using the PLS loadings plots. Descriptors were manually removed if they had minimal influence on the first two components.

To determine if the model could have occurred by chance Y-randomization trials of the training set matrix were performed within SIMCA. A total of 700 training sets were randomly generated by scrambling the logK values, and the models were rebuilt. In all cases, the randomized models had negative q^2 and r^2 values no greater than 0.1, implying that the real PLS models are considerably better than random. Plots of observed versus predicted values were produced in JMP 5.1.[29]

The distance of a compound j to the training set space was assessed using the components that define the PLS models. The distance of a compound to the training set is computed as sum of the absolute products of the *Z* scaled and centered descriptors (*c*) multiplied by their corresponding PLS weights (*w*), over each of the *N* principal components of the model in question (eq 4). The distance is then normalized by dividing the average distance of the dataset in question by the average of the training set compounds' distances. This means a test or validation set with an average distance less than 1 lies within the model set space and far outside if much greater than 1.

$$\mathrm{dist_j} = \sum_{K=1}^{N} \mathrm{abs}\left(\sum_{j=1}^{Z} c_{ij} w_{ik}\right) \tag{4}$$

The relationship between the distance and the errors was assessed by binning the test set into quartiles before assessing whether the RMSE increased with increasing distance. Since the absolute errors of bin are not normally distributed (the absolute errors follow an exponential distribution compared to the normally distributed errors), it is not possible to assess the statistical significance of the mean values of the bins using parametric statistics such as ANOVA. Thus, a nonparametric test (Kruskal-Wallis ANOVA) was employed.[30] This test assesses whether the frequency of compounds found above the median of the dataset increases as one moves from the first to the fourth quartile. The chi statistic is then used to quantify the statistical significance of the relationship.

4.3. Validation Statistics. The coefficient of determination (r_0^2) is used as the primary measure to estimate the fraction of the total sum of square variance explained by the model (i.e., correlation to the line of unity). It has been rearranged from the more commonly used equation so that it can be expressed in terms of the root-mean-square error and the variance of the observed value.[31] Often, predictions of a dataset may give rise to a low r_0^2 even though the model may be useful to rank order compounds in that set. In such cases, the slope and intercept of the line of best fit deviate from 1 and 0, respectively, and it is useful to evaluate the Pearson product moment correlation coefficient (r^2). The root-mean-square error in prediction (RMSE) is also calculated as this can be directly related back to the known experimental errors. The mean error in prediction (ME) indicates whether a model experiences an over- or under-prediction. The cross-validated q^2 is reported, calculated in SIMCA-P10 by generating seven different models, each by leaving 1/7 of the data out each time. Finally, the maximum possible r_0^2/r^2 was estimated by substituting the experimental error (∼0.25 log units) into eq 5 rather than the in-silico RMSE.

The r_0^2, r^2, RMSE, and ME are calculated using eqs 5−8, respectively, where *x* and *y* are predicted and observed values of the biological response, *n* is the number of compounds, and σY is the standard deviation of the *y* values. The normal probability plots of the quoted statistical values have been visually assessed to ensure their validity.

$$r_o^2 = 1 - \frac{n}{n-1}\left(\frac{\mathrm{RMSE}^2}{\sigma Y^2}\right) \tag{5}$$

$$r^2 = \left(\frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2]\cdot[n\sum y^2 - (\sum y)^2]}}\right)^2 \tag{6}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum (y-x)^2}{n}} \tag{7}$$

$$\mathrm{ME} = \frac{\sum (y-x)}{n} \tag{8}$$

Supporting Information Available: A table with the observed and predicted values for the literature model generated herein. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Grandison, M. K.; Boudinot, F. D. Age-related changes in protein binding of drugs. Implications for therapy. *Clin. Pharmacokinet.* **2000**, *38*, 271−290.

(2) Benet, L. Z.; Hoener, B. A. Changes in plasma protein binding have little clinical relevance. *Clin. Pharmacol. Ther.* **2002**, *71*, 115−121.

(3) Klippert, P.; Borm, P.; Noordhoek, J. Prediction of intestinal first-pass effect of phenacetin in the rat from enzyme kinetic data − correlation with in vivo data using mucosal blood flow. *Biochem. Pharmacol Pharmacol.* **1982**, *31*, 2545−2548

(4) Oie, S.; Tozer, T. N. Effect of altered plasma protein binding on apparent volume of distribution. *J. Pharm. Sci.* **1979**, *68*,1203−1205.

(5) Kratochwil, N. A.; Huber, W.; Muller, F.; Kansy, M.; Gerber, P. R. Plasma protein binding of drugs. A new approach. *Biochem. Pharmacol.* **2002**, *64*, 1355−1374.

(6) Colmenarejo, G.; Alvarez-Pedraglio, A.; Lavandera, J. L. Cheminformatic models to predict binding affinities to human serum albumin. *J. Med. Chem.* **2001**, *44*, 4370−4378.

(7) Lobell, M.; Sivarajah, V. In silico prediction of aqueous solubility, human plasma protein binding and volume of distribution of compounds from calculated pK_a and AlogP98 values. *Mol. Diversity* **2003**, *7*, 69−87.

(8) Yamakazi, K.; Kanaoka, M. Computational Prediction of plasma protein binding Percent of diverse pharmaceutical compounds. *J. Pharm. Sci.* **2004**, *93*, 1480−1494.

(9) Saiakhov, R. D.; Stefan, L. R.; Klopman, G. Multiple computer-automated structure evaluation model of plasma protein binding affinity of diverse drugs. *Perspect. Drug Discovery* **2000**, *19*, 133−155.

(10) Austin, R. P.; Barton, P.; Mohmed, S.; Riley, R. J. The binding of drugs to hepatocytes and its relationship to physicochemical properties. *Drug Metab. Disp.* **2005**, *33*, 419−425.

(11) Singh, S. S.; Mehta, J. Measurement of drug-protein binding by immobilized human serum albumin-HPLC and comparison with ultrafiltration. *J. Chromat. B* **2006**, *834*, 108−116.
(12) Valko, K.; Nunhuck, S.; Bevan, C.; Abraham, M. H.; Reynolds, D. P. Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci.* **2003**, *92*, 2236−2248.
(13) Gleeson, M. P.; Waters, N. J.; Paine, S. W.; Davis, A. M. In Silico Human and rat V_{ss} quantitative structure-activity relationship models. *J. Med. Chem.* **2006**, *49*, 1953−1963.
(14) Stouch, T. R.; Kenyon, J. R.; Johnson, S. R.; Chen, X. Q.; Doweyko, A.; Li, Y. J. In silico ADME/Tox: why models fail. *Comput.-Aided Mol. Des.* **2003**, *17*, 83−92.
(15) Zlotos, G.; Bucker, A.; Kinzig-Schippers, M.; Sorgel, F.; Holzgrabe, U. Plasma protein binding of gyrase inhibitors. *J. Pharm. Sci.* **1998**, *87*, 215−220.
(16) Advanced Chemistry Development, Inc., 110 Yonge Street, 14th Floor, Toronto, Ontario, M5C 1T4, Canada. http://www.acdlabs.com.
(17) %deprotonated = $[100/(10^{(pKa\ acid-7.4)+1)}]$; %protonated = $[100/(10^{(7.4-pKa\ base)+1)}]$; %zwitterion = smaller of [%deprotonated/2 or %protonated/2], %basic = [%protonated − %zwitterion], %acidic = [%deprotonated − %zwitterion], %neutral = 100 − %acid − %basic − %zwitterion.
(18) Ertl, P.; Rohde, B.; Selzer, P. J. Fast calculation of molecular polar surface area as a sum of fragment based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.* **2000**, *43*, 3714−3717.
(19) Abraham, M. H.; Chadha, H. S.; Mitchell, R. C. Hydrogen bonding. 33. Factors that influence the distribution of solutes between blood and brain. *J. Pharm. Sci.* **1994**, *83*, 1257−1268.
(20) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional group contributions to drug-receptor interactions, *J. Med. Chem.* **1984**, *27*, 1648−1657.
(21) Daylight Chemical Information Systems, Inc., 120 Vantis, Suite 550, Aliso Viejo, CA 92656. http://www.daylight.com.
(22) The ionization state of a molecule was defined based on the presence or absence of acidic or basic functional groups which are predominantly in the ionized form at pH 7.4. Neutral molecules had none of the in-house defined functional groups, while zwitterions had at least one acidic and one basic group present.
(23) Statistica 7.1, Statsoft Inc, 2300 East Tulsa, OK 74104, U.S.A, http://www.statsoft.com.
(24) Höskuldsson, A. Prediction Methods in Science and Technology, Thor Publishing: Copenhagen, Denmark, 1996.
(25) Wold, S.; Albano, C.; Dunn, W. J.; Edlund, U.; Esbensen, K.; Geladi, P.; Hellberg, S.; Johansson, E.; Lindberg, W.; Sjöström, M. Multivariate Data Analysis in Chemistry, In *Chemometrics: Mathematics and Statistics in Chemistry*; B. R. Kowalski, Ed.; D. Reidel Publishing Company: Dordrecht, Holland, 1984.
(26) Wold, S.; Eriksson, L.; Sjöström, M. PLS in Chemistry, Encyclopedia of Computational Chemistry, Wiley: New York, 2000.
(27) GOLPE: Multivariate Infometric Analysis Srl., Viale dei Castagni 16, Perugia, Italy.
(28) SIMCA-P 10, Umetrics, Tvistevägen 48, Box 7960, SE-907 19 Umeå, Sweden.
(29) JMP 6.0, JMP Europe, SAS Institute, Henley Road, Medmenham, Marlow, SL7 2EB UK, http://www.jmp.com.
(30) Statistica System Reference, Statsoft Inc, 2300 East Tulsa, OK74104, USA, http://www.statsoft.com.
(31) The equation has been arranged from the commonly used equation so that it can be expressed in terms of the root mean square error and the variance of the observed value. The line of unity $r^2 = 1 - [\sum(y_{(obs)} - y_{(pred)})^2]/[\sum(y_{(obs)} - y_{(mean)})^2]$ and since RMSE = $[(\sum(y_{(obs)} - y_{(pred)})^2]/(N - 1)]^{1/2}$ and $\sigma Y = [(\sum(y_{(obs)} - y_{(mean)})/N]^{1/2}$, one can rearrange the line of unity r^2 to give eq 5.
(32) The one-way ANOVA results and the two-way ANOVA results associated with Figure 1 are statistically significant above the 95% confidence level, indicating both ionization state and logP have an independent impact on the extent of protein binding. Two-tailed T-tests between the different ionization states indicate the logK values of neutrals and zwitterions are equivalent but those for acids and bases are significantly different. The same tests indicate the affect of increasing logP, as given by the <3, 3−5 & >5 bins, are also statistically significantly different.

JM060981B