**Relating Caco-2 permeability to molecular properties using block relevance analysis**



MedChemComm

*Volume 6, Issue 4, 2015*

*DOI: 10.1039/C4MD00470A*

Journal Name

RSCPublishing

COMMUNICATION

# Relating Caco-2 permeability to molecular properties using block relevance analysis

Cite this: DOI: 10.1039/xoxxooooox

T. Potter[a], G. Ermondi[b], G. Newbury[a] and G. Caron[b]

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxooooox

www.rsc.org/

**Here we describe a new approach for facilitating a mechanistic understanding of high throughput Caco-2 permeability data. A large, uniform set of permeability data is reported, generated under two pH conditions. We found that this data could be successfully modelled, and that a mechanistic understanding of apparent permeability could be gained.**

## Introduction

The Caco-2 cell model is widely used in drug discovery and development for the determination of permeability characteristics of potential drug candidates. Reports have demonstrated its utility in the prediction of oral absorption of drugs in man based on permeability across Caco-2 monolayers[1].

Several previous studies of the modelling of permeability across Caco-2 monolayers have been reported[2,3,4,5,6,7,8,9]. The main purpose of such modelling was to successfully predict the screening result, and in some cases to act as a pre-screen prior to the measurement. Typically, researchers found that physicochemical properties are essential to define Caco-2 permeability, but often the relative contribution of each property was not explored.

High throughput (HT) Caco-2 methods are now available for testing of high numbers of compounds in discovery phases, often replacing PAMPA as the screening method of choice. We have developed a HT Caco-2 permeability assay capable of running 96 samples in a single experiment, using simultaneous quantification of samples from A-B and B-A directions. A large dataset of apparent permeability coefficient ($P_{app}$) values have been generated, in order to validate the assay, and to compare against reported human intestinal absorption measurements[1].

A critical analysis of the relative contribution of molecular properties governing the permeability process could be used as a tool for comparing methods set-up in different labs, as well as for helping to guide medicinal chemistry programs. For the aforementioned reasons we were interested in mining this dataset to understand further the fundamental drivers and their relative contribution of permeability in HT Caco-2 assays. Following recent guidance for QSAR modelling[10], we set out to find a mechanistic interpretation, with clear relation to physicochemical properties.

Block Relevance (BR) analysis is a new tool that produces an easy-to-read mechanistic interpretation of PLS models based on VolSurf+ (VS+) descriptors.[11, 12, 13] The basic concept of VolSurf is to extract the information present in 3D molecular field maps into a few quantitative numerical descriptors which are easy to understand and to interpret.

We have successfully applied the technique to distinguish chromatographic indexes[11,12] and to characterize the dominant effect of Hydrogen Bond Donor (HBD) solute properties in the difference between log P measurements in two different systems (log $P_{oct\text{-}tol}$)[13]. BR analysis mandates the organization of the VS+ descriptors into six blocks (namely, Size, Water, DRY, N1, O and Others) which enable a straightforward understanding of the investigated phenomenon, in this case cell-based permeability. We sought to investigate from BR outputs, whether permeability across Caco-2 monolayers is driven predominantly by a compounds hydrophobicity or hydrogen bonding properties.

This paper reports a large and uniform set of original apical to basolateral permeability data across Caco-2 monolayers and highlights which intermolecular forces drive permeability. These results could be used to aid medicinal chemistry design, and also provide an additional validation criteria to evaluate HT Caco-2 permeability assays, based on the balance of the intermolecular forces governing the system.

## Results and Discussion

$P_{app}$ was determined in Caco-2 cells from equation 1, where dQ/dt is the rate of permeation of the drug across the cells, Co is the donor

compartment concentration at time zero and A is the area of the cell monolayer.

$$\mathbf{P_{app}} = \left(\frac{\mathbf{dQ/dt}}{\mathbf{C_0 \times A}}\right)$$

Equation 1.

Two different pH conditions for the donor compartment (6.5 and 7.4) were used, producing two sets of data; $P_{app}$ 6.5/7.4 (n = 85) and $P_{app}$ 7.4/7.4 (n = 54), respectively. These represent two of the most commonly screened conditions for Caco-2 permeability. An acidic apical compartment can be used to model the upper small intestine, for which the pH6.5/7.4 screening condition is employed. However many researchers prefer to use pH7.4 in both chambers to eliminate the effect of a pH gradient. Test compound permeability was assessed in duplicate, and the mean $P_{app}$ values, along with standard deviations are reported in the supporting information.

The dataset was checked for its chemical diversity through the analysis of some physicochemical properties (see supporting information). Results showed that a broad range of physicochemical properties was covered.

The $P_{app}$ 6.5/7.4 dataset was split in a training (n = 54) and a test set (n = 31). The training set was designed to include drugs for which both $P_{app}$ 6.5/7.4 and $P_{app}$ 7.4/7.4 were available. We verified the relationship between the two series of data ($P_{app}$ 6.5/7.4 and $P_{app}$ 7.4/7.4), and found a very good correlation ($R^2$ = 0.93, not shown), with no significant outliers.

Before proceeding with modeling we made an assumption regarding molecular flexibility. When VS+ processes the data, it associates each compound to the lipophilicity value of an "average" conformer built internally by an ad-hoc algorithm. In general terms this is a suitable protocol because one can assume that the "average" conformer represents all conformers energetically accessible. However this assumption no longer holds when the molecule in question has strong propensity to form intra-molecular hydrogen bonds, since the molecule is then forced into a specific conformation. In this case a very different profile of VolSurf+ descriptors could be obtained from the conformers without intra-molecular hydrogen bonds.[13] In the process of crossing the cell membrane, the molecules experience a wide variety of environments (membrane interaction, diffusion, flip-flop mechanism).[14, 15] Conformation may vary greatly during the membrane passage and thus the use of an average conformation would seem reasonable.

Finally we made some preliminary tests that enabled the identification of three outliers: acarbose, digoxin and adefovir. Acarbose and digoxin both possess a tri-saccharide chain containing multiple hydrogen bond donors and acceptors and are hence outliers in terms of these descriptors. Adefovir could have parameterisation issues in the GRID force field applied by VS+ (see Experimental Section). All three outliers were removed from the study.

## Modeling log $P_{app}$ (6.5/7.4)

Experimental log $P_{app}$ values of the compounds belonging to the training set (n = 54) were imported into VS+ as response variables (Y) and a relation between Y and the 82 VS+ descriptors (X) was sought using the PLS algorithm implemented in the software. A model was found (Table 1) and the correlation between calculated vs experimental values is shown in Fig. 1 (filled circles).

| Activity | N | Validation | LVs | $R^2$ | $Q^2$ | RMSE |
|---|---|---|---|---|---|---|
| log $P_{app}$ 6.5/7.4 | 54 | LOO | 3 | 0.72 | 0.50 | 0.61 |
| | 54 | RG | 3 | 0.71 | 0.49 | 0.62 |
| log $P_{app}$ 7.4/7.4 | 54 | LOO | 3 | 0.79 | 0.57 | 0.56 |
| | 54 | RG | 3 | 0.78 | 0.56 | 0.57 |
| log D7.4 | 54 | LOO | 3 | 0.70 | 0.52 | 0.80 |
| | 54 | RG | 3 | 0-70 | 0.50 | 0.80 |

Table 1. PLS statistical results

(N = number of observations, R2 = cumulative determination coefficient, Q2 = cross-validated correlation coefficient, LV = number of latent variables, RMSE = root mean square error, LOO = Leave One Out, RG = Random Group))

The validation of the models was firstly performed by means of an internal procedure (see experimental section). Satisfactory statistical results were obtained (Table 1).

Following recently published QSAR guidelines[10] we also performed an external validation of the PLS model using the test set. Predictions are shown in Fig. 1 (empty circles) and support the statistical stability of the model.

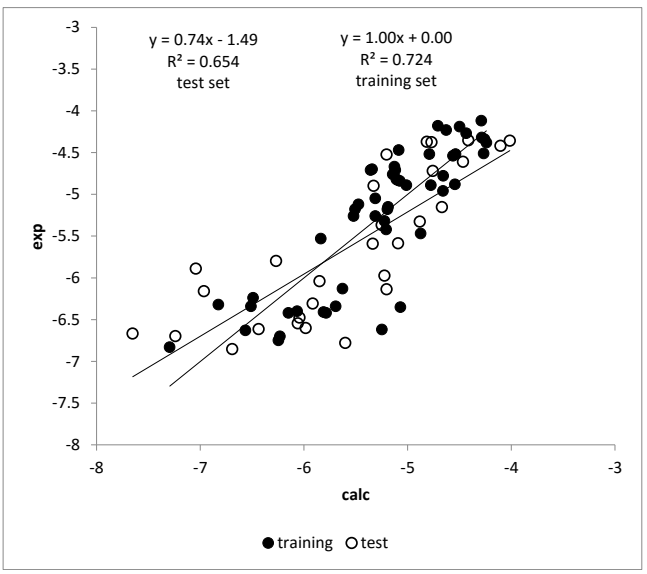


Figure 1. Experimental log $P_{app}$ (6.5/7.4) of the drugs belonging to the training set versus the predicted values (filled circles) and experimental log $P_{app}$ (6.5/7.4) of the drugs belonging to the test set versus the predicted values (empty circles).

## Modeling log $P_{app}$ (7.4/7.4)

The same procedure described above for log $P_{app}$ (6.5/7.4) was also performed for log $P_{app}$ (7.4/7.4). The validation of the models was performed by means of the internal procedures alone (LOO, RG and dataset with a randomized Y order). Satisfactory statistical results were obtained (Table 1), slightly better than those obtained when modeling log $P_{app}$ (6.5/7.4).

## Interpretation of the PLS models by Block Relevance (BR) analysis

A mechanistic interpretation based on the nature of the interaction of the solute with the environment, represented by some tailored probes, 'blocks' is defined by the GRID methodology[16, 17, 18]. Their significance is summarized in Fig. 2, as proposed in previous studies[12].

| **Size** *Volume and surface* | **OH2** *Molecular polarity* | **DRY** *Hydro-phobicity* | **O** *H-Bond donor properties* | **N1** *H-Bond acceptor properties* | **Other** *Polarity unbalance* |
|---|---|---|---|---|---|
| molecular hydrophobic interaction with the system mainly of entropic nature | the interaction of the polar regions of the solute with the system | local interactions between apolar regions of the solute and the system | specific HB interactions between solute and system | specific HB interactions between solute and system | difference in interactions of solutes and system due to different location of polar and apolar regions |

Figure 2. Block definition: the name of the block is in bold, the solute's property described by the block in *italics*. The color code is maintained throughout the study.

The BR analysis results are graphically reported in Fig.3A and Fig.3B.

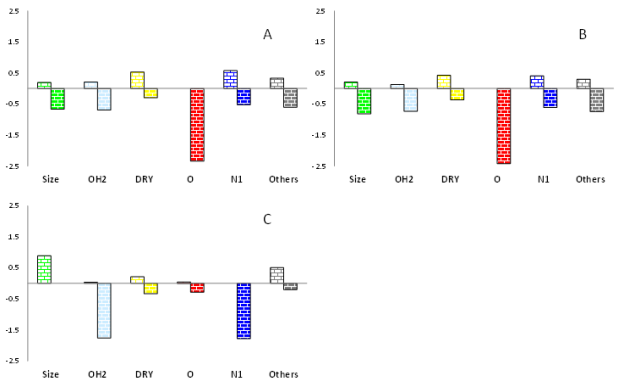


Figure 3. BR analysis graphical output: A) log $P_{app}$ (6.5/7.4); B) log $P_{app}$ (7.4/7.4); C) calculated log Doct 7.4. The blocks' definition is reported in Figure 2.

Blocks with a positive weighting indicate how much the property increases permeability, whereas those with negative weighting (e.g. red block) indicate how much the property decreases permeability. The presence of positive and negative components for most blocks is partly due to noise and partly due to the inter-correlation of descriptors.

Figure 3A (log $P_{app}$ 6.5/7.4) and 3B (log $P_{app}$ 7.4/7.4) outline the major role (about 35% of the weight of all blocks) played by HBD solute properties (red bars) to govern Caco-2 permeability. The role of HBA solute groups (blue bars) by contrast is modest (about 15%) and similar to all remaining blocks.

Very little difference was observed between the two data sets in terms of the output from BR analysis. A possible reason for this is the dataset does not contain a large number of compounds that we would expect to have different charge states at pH 6.5 versus 7.4.

It is often reported in the literature that permeability can be modeled by log D7.4oct, but this analysis would seem to suggest that logD alone is a poor surrogate. To investigate further, we calculated log D7.4oct for all the drugs, and plotted against log $P_{app}$ 6.5/7.4, Figure 4.

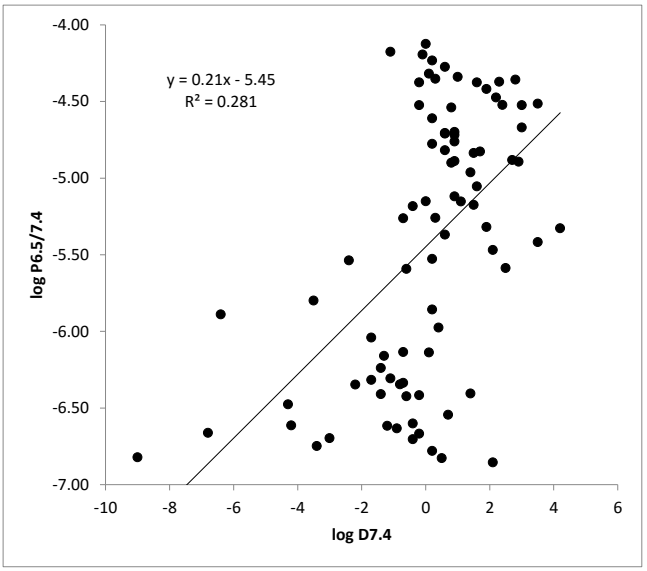


Figure 4. The relationship between log $P_{app}$ (6.5/7.4); and calculated log Doct 7.4.

The weak correlation observed suggests that the balance of intermolecular forces that govern log D7.4oct and log $P_{app}$ 6.5/7.4 is very different. To confirm this hypothesis we again used BR analysis.

The log D7.4oct values of the training set were imported into VS+ as response variables (Y) and a relation between Y and the 82 VS+ descriptors (X) was sought using the PLS algorithm as previously described. Statistics were good (see table 1) and thus we submitted PLS output to BR analysis. The graphical output is shown in Fig. 3C and shows how the balance of intermolecular interactions governing log Doct 7.4 is completely different from that governing log $P_{app}$. (Fig. 3A and 3B) This supports the finding that lipophilicity is not the dominant factor for understanding log $P_{app}$. In fact, a large drug size (green block) generally causess an increase in log Doct 7.4 (Fig. 3C, positive sign) but a decrease in log $P_{app}$ (Fig. 3A and 3B, negative sign). Moreover HBA solute properties (blue bars) are important for logD7.4 but not for log Papp as previously discussed.

## Conclusions

This analysis has shown that the most important factors for understanding the physicochemical drivers for HT Caco-2 permeability are hydrogen bond donor properties. No consideration has been given in this analysis to the contribution of active transport mechanisms and the effect they may have on $P_{app}$. Clearly such factors are of great importance when describing permeability in Caco-2 cells, and compounds which display significant active efflux will inevitably have a low apical to basolateral $P_{app}$ value which is not fully explained by an analysis of bulk physicochemical properties such as this. These interactions may, in part be accounted for in the analysis, but we have

not attempted to include sufficient compounds in the dataset to model active efflux. The conclusions should therefore be interpreted as a bulk property basis, and not in a purely predictive fashion. Indeed, we would not expect compounds with specific interactions to be well predicted by this model. However, we would expect the same principles to apply to the passive aspect of such compounds permeability.

We now intend to extend this analysis to assess the main drivers for human intestinal absorption, and to verify (or otherwise) that the same descriptors and directionalities are driving this phenomenon as Caco-2 permeability.

## Experimental

All reagents were purchased from Sigma-Aldrich unless otherwise stated. Hanks Balanced Salt Solution (HBSS) buffer was supplemented with 25mM HEPES and 4.45 mM glucose and the pH adjusted to 7.4. HBSS buffer was supplemented with 10 mM MES and 4.45mM glucose, and the pH adjusted to 6.5.

Caco-2 cells (ATCC) were seeded onto Millicell 96 well plates (Millipore, MA, USA) at 1 x $10^5$ cells/$cm^2$. The cells were cultured in DMEM and media was changed every two or three days for 20 days for confluent cell monolayer formation. Cell culture and assay incubations were carried out at 37°C, 5% $CO_2$ with a relative humidity of 95%. On the day of the experiment, the monolayers were prepared by rinsing both apical and basolateral surfaces twice with HBSS at the desired pH. Cells were incubated with HBSS at the desired pH in both apical and basolateral compartments for 40 minutes to stabilise physiological parameters.

The dosing solutions were prepared by diluting test compound with assay buffer to give a final test compound concentration of 10 µM (final DMSO concentration of 1 % v/v). The fluorescent integrity marker lucifer yellow was also included in the dosing solution. Analytical standards were prepared from test compound DMSO dilutions and transferred to buffer, maintaining a 1 % v/v DMSO concentration. For assessment of A-B permeability, HBSS was removed from the apical compartment and replaced with test compound dosing solution. The apical compartment insert was then placed into a companion plate containing fresh buffer (containing 1 % v/v DMSO). At 120 minutes the apical compartment inserts and the companion plates were separated and apical and basolateral samples diluted for analysis.

Test compounds were quantified by LC/MS/MS analysis using an 8 point calibration with appropriate dilution of the samples. The integrity of the monolayer throughout the experiment was checked by monitoring lucifer yellow permeation using fluorimetric analysis. If lucifer yellow permeation was found to be above pre-defined thresholds, the result was rejected and repeated.

VS+ models were built by submitting the SMILES codes of the compounds to VS+ (version 1.0.7, http://www.moldiscovery.com) using default settings and four probes (OH2, DRY N1 and O probes that mimic respectively water, hydrophobic, HBA and HBD properties of the environment). PCA and PLS tools implemented in VS+ were used. BR analysis was performed as described elsewhere.[11, 12, 13]

$pK_a$ and log D7.4 calculations were performed with MoKA (v.2.5.4, www.moldiscovery.com).

Processing was done on a two 8 cores Xeon E5 at 3.3GHz CPUs and 128GB of RAM.

## Notes and references

[a] Cyprotex Discovery Limited, 15 Beech Lane, Macclesfield, SK10 2DR, UK.

[b] CASSMedChem research group, Molecular Biotechnology and Health Sciences Dept., University of Turin.

Electronic Supplementary Information (ESI) available: Full table of measured Caco-2 permeability used in the modelling. See DOI: 10.1039/c000000x/

1 Y. Zhao, J. Le, M. Abraham, A. Hersey, P. Eddershaw, C. Luscombe, D, Boutina, G. Beck, B. Sherborne, I. Cooper, J. Platts, J. Pharm. Sci. 2001, 90, 749.

2 P. Paixão, L. Gouveia, J. Morais, European Journal of Pharmaceutical Sciences, 2010, 41, 107.

3 T. Hou, W. Zhang, K. Xia, X. Qiao, X. Xu, J. Chem. Inf. Comput. Sci, 2004, 44, 1585.

4 A. Nordqvist, J. Nilsson, T. Lindmark, A. Eriksson, P. Garberg, M. Kihlén, QSAR & Combinatorial Science. 2004, 23, 303.

5 Y. Rojas-Aguirre, J. L. Medina-Franco, Molecular Diversity. 2014, 18, 599.

6 H. Pham-The, T. Garrigues M. Bermejo, I. González-Álvarez, M. Monteagudo, M. Cabrera-Pérez, Mol Pharm, 2013, 10, 2445.

7 Mar F. Meng, W. Xu, J Mol Model, 2013, 19, 991.

8 O. Santos-Filho A. Hopfinger. J Pharm Sci. 2008, 97, 566.

9 A. Di Fenza, G. Alagona, C. Ghio, R. Leonardi, A. Giolitti, A. Madami, J Comput Aided Mol Des. 2007, 21, 207.

10 A. Cherkasov, A, E. Muratov, D. Fourches, A. Varnek, I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. Martin, R. Todeschini, V. Consonni, V. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, J. Gasteiger, A. Richard, A. Tropsha, J. Med. Chem. 2014, 57, 4977.

11 G. Ermondi, G. Caron, J. Chrom. A. 2012, 1252, 84.

12 G. Caron, M. Vallaro, G. Ermondi, MedChemComm, 2013, 4, 1376.

13 G. Ermondi, A. Visconti, R. Esposito, G. Caron, Eur. J. Pharm. Sci, 2014, 53, 50.

14 S. Kraemer, D. Lombardi, A. Primirac, A. Thomae, H. Wunderli-Allenspach, 2009, 6, 1900.

15 J. Alakoskela, P. Vitovic, P. Kinnunen, ChemMedChem, 2009, 4, 1224.

16 P. Goodford, J. Med. Chem., 1985, 28, 849.

17 D. Boobbyer, P. Goodford, P. McWhinnie, R. Wade, J. Med. Chem, 1989, 32, 1083.

18 R. Wade, P. Goodford, J. Med. Chem.; 1993, 36, 148.

 