The apparent permeabilities of Caco-2 cells to marketed drugs: magnitude, and independence from both biophysical properties and endogenite similarities

[1,2,3]Steve O'Hagan & [1,2,3,*]Douglas B. Kell

[1]School of Chemistry, [2]The Manchester Institute of Biotechnology, and [3]Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), The University of Manchester, 131, Princess St, Manchester M1 7DN, United Kingdom.

sohagan@manchester.ac.uk, dbk@manchester.ac.uk

*corresponding author: Tel 0044 161 306 4492 dbk@manchester.ac.uk http://dbkgroup.org @dbkell

Abstract

We bring together fifteen, nonredundant, tabulated collections (amounting to 696 separate measurements) of the apparent permeability (P_{app}) of Caco-2 cells to marketed drugs. While in some cases there are some significant interlaboratory disparities, most are quite minor. Most drugs are not especially permeable through Caco-2 cells, with the median P_{app} value being some 16 . 10^{-6} $cm.s^{-1}$. This value is considerably lower than those (1310 and 230 . 10^{-6} $cm.s^{-1}$) recently used in some recent simulations that purported to show that P_{app} values were too great to be transporter-mediated only. While these values are outliers, all values, and especially the comparatively low values normally observed, are entirely consistent with transporter-only mediated uptake, with no need to invoke phospholipid bilayer diffusion. The apparent permeability of Caco-2 cells to marketed drugs is poorly correlated with either simple biophysical properties, the extent of molecular similarity to endogenous metabolites (endogenites), or any specific substructural properties. In particular, the octanol:water partition coefficient, log P, shows negligible correlation with Caco-2 permeability. The data are best explained on the basis that most drugs enter (and exit) Caco-2 cells via a multiplicity of transporters of comparatively weak specificity.

Keywords

Caco-2 cells, Cheminformatics, Facilitated diffusion/transport, Mathematical models, Oral absorption, Permeability, Transcellular transport, Transporter-mediated uptake, Transporters

Introduction

Most pharmaceutical drugs, and all oral ones, must necessarily cross at least one cell membrane to act. Understanding how this transport is effected remains a major challenge (Kell & Oliver 2014). We have brought together considerable published evidence (e.g. (Dobson & Kell 2008; Kell 2013; Kell 2015; Kell et al. 2013; Kell et al. 2011; Kell & Oliver 2014)) that suggests that (in contrast to the general textbook belief, e.g. (Avdeef 2012; Cao et al. 2006; Krogsgaard-Larsen et al. 1996; van De Waterbeemd & Testa 2009)) small molecule drugs 'hitchhike' on the many protein transporters (Kell 2013; Kell & Goodacre 2014; Sahoo et al. 2014; Thiele et al. 2013) that are part of normal intermediary metabolism. These transporters may be identified via experiments where gene expression levels are manipulated systematically as independent variables (Giacomini et al. 2010; Han et al. 2015; Kell & Oliver 2014; Lanthaler et al. 2011; Winter et al. 2014). A number of recent books summarise the importance of protein transport to drug disposition (Bhardwaj et al. 2008; Ecker & Chiba 2009; Fromm & Kim 2011; Ishikawa et al. 2013; Sugiyama & Steffansen 2013; You & Morris 2014).

Caco-2 cells (e.g. (Artursson et al. 2001; Awortwe et al. 2014; Balimane & Chong 2005; Fearn & Hirst 2006; Feng et al. 2014; Hidalgo et al. 1989; Sarmento et al. 2012; Sun et al. 2008; van Breemen & Li 2005; Volpe 2011)) are an epithelial cell line that has become a *de facto* standard in studies of pharmaceutical drug transport. They form a more or less (and otherwise) impermeable layer that is polarised, in the sense of having 'apical' and 'basolateral' faces in which transporters are differentially expressed. They express hundreds of transporters (Anderle et al. 2004; Hayeshi et al. 2008; Landowski et al. 2004; Pshezhetsky et al. 2007; Sun et al. 2002), and (although far from perfect (Hilgendorf et al. 2007)) they have significant predictive power as to the fraction of oral dose absorbed in humans (e.g. (Marino et al. 2005; Rubas et al. 1996)).

It is thus of general interest to understand the kinds of apparent permeability (P_{app}) rates for different drug molecules that Caco-2 cells can sustain. Although there are undoubtedly larger databases in-house in commercial and other enterprises, we have sought to bring together what we can of published data to determine the kinds of permeability values that Caco-2 cells can sustain, and what might determine that. We recognise that many factors can affect a specific measurement, e.g. the seeding density, age of the cells, pH and so on. An interlaboratory comparison (Hayeshi et al. 2008) indicated that while on occasion measurements could vary by more than an order of magnitude, overall the groupings were normally reasonably tight (say within a factor of 2-5).

The question of P_{app} values in Caco-2 cells has been brought into sharper focus by a recent article (Matsson et al. 2015a; Matsson et al. 2015b) that claimed unusually high rates for verapamil and propranolol, based on measurements a specific earlier article (Avdeef et al. 2005) in which stirring had been performed at a massive rate (and one not used in any equivalent transporter kinetic measurements). We indicated that these values were major outliers (by one or even two orders of magnitude) (Mendes et al. 2015), but did not pursue the question of typical values of P_{app} for other drugs. This is the focus of what we do here.

Methods

Data were extracted manually from tables in the papers stated, and compiled as an Excel sheet. Typical biophysical descriptors were added using the RDKit module (Riniker & Landrum 2013) of KNIME (Berthold et al. 2008; Mazanetz et al. 2012; Saubern et al. 2011) (www.knime.org/), essentially as described (O'Hagan & Kell 2015a; O'Hagan & Kell 2015b; O'Hagan et al. 2015). For one experiment we used the CDK-KNIME nodes (Beisken et al. 2013).

Results

We have selected a set of 15 studies (indicated in the legend to Figure 1) for our analysis. Based on the list of FDA-approved drugs that we downloaded (as before (O'Hagan & Kell 2015b; O'Hagan et al. 2015)) from DrugBank (http://drugbank.ca) (Law et al. 2014), we compiled from these a non-redundant set of measurements of the apparent permeability (P_{app}, that are commonly given in units of $cm.s^{-1}$). Although there are older papers, we have started with the compilation of Hou and colleagues (Hou et al. 2004). Our method for avoiding redundancy in later compilations was not to include a separate measurement if the numbers given were identical to those in Hou (Hou et al. 2004) (or any other later papers) to at least 1 decimal place. We ignore any efflux transporters, since the evidence (that we show later) is that their influence on these measurements is fairly small (Lin et al. 2011). We incorporated two values from the review of Marino and colleagues (Marino et al. 2005), one from lower throughput 24-well plates, one from a 96-well assay.

Where data were available for bidirectional assays, e.g. (Hayeshi et al. 2008; Skolnik et al. 2010) they are given just for the A → B direction. In the case of the interlaboratory comparison (Hayeshi et al. 2008), we used solely 'batch 1' data, while in the work of Lin et al. (Lin et al. 2011) efflux inhibitors were sometimes present, as noted below. The entire dataset is given as an Excel sheet as Supplemental Table S1, and consists of 696 separate measurements. As indicated in Methods, we used KNIME to append some simple biophysical descriptors.

Figure 1A shows all of the data, with those studies finding rates above 100.10^{-6} $cm.s^{-1}$ labelled with the study number. Of the 21 measurements that have this property, no fewer than 9 (labelled in red) are from a study (Avdeef et al. 2005) of Avdeef and colleagues. The largest values (Avdeef et al. 2005) were observed at very high values of stirring rates (700 rpm), and these in particular contained a great many outliers. The implication is that these increases at exceptionally high stirring rates were due to unstirred layer effects, although it is hard to see their relevance to *in vivo* drug absorption where no such stirring is occurring. Mannitol is sometimes used as a membrane-impermeant control, taken to pass via a paracellular route. This said, mannitol controls did not always have the lowest values, and inulin (Marino et al. 2005) or EDTA (Lin et al. 2011) may be better. Although it was stated (Avdeef et al. 2005) that mannitol transport rates were 'normal', it is unclear why they do not change with stirring rates (or whether they do), so it is not entirely certain whether the epithelial layer remained intact, especially at some of the highest stirring rates employed. For these and other reasons, and especially given the strongly outlying nature of the measurements, we have decided for the rest of the analysis to exclude the data from (Avdeef et al. 2005), resulting in an overall dataset of 680 separate measurements as shown in Fig 1B. A cumulative plot and smoothed histogram of the data (Fig 1C) shows that the most abundant values for P_{app} are in the range 3 to 4 . 10^{-6} $cm.s^{-1}$, and with a median value of ca 16 . 10^{-6} $cm.s^{-1}$. Obviously these values are considerably lower than those discussed in (Matsson et al. 2015a; Matsson et al. 2015b), and indicate (Mendes et al. 2015) that typical transporter kinetic parameters and expression levels are entirely adequate to account alone for cellular drug uptake, as proposed (Dobson et al. 2009a; Dobson & Kell 2008; Kell 2013; Kell 2015; Kell & Dobson 2009; Kell et al. 2013; Kell et al. 2011; Kell & Goodacre 2014; Kell & Oliver 2014; Kell et al. 2015).

Figure 2 illustrates another feature of the data. Here we took the tabulated data of Lin, Skolnik and colleagues (Lin et al. 2011) that used a variety of efflux inhibitors. A comparison showed that no very substantial (order-of-magnitude) differences in uptake were observed (Fig 2), such that the typical 'low' values of P_{app} cannot realistically be ascribed to a major role of efflux pumps.

Lack of relationship between Caco-2 permeability values and simple biophysical properties of drugs

If unstirred layer effects and pure diffusion (as opposed to transporter-based enzyme kinetics) were significant in Caco-2 permeability, one might suppose that permeability values should depend significantly upon the molecular mass of the drug involved. However, Fig 3A shows that this is not the case, as the line of best fit has a slope of only -0.04X and a value for r^2 of just 0.069. In a similar vein, despite a widespread view that transport rates should depend on log P, Fig 3B shows that even when the Caco-2 permeabilities are plotted in log space, the r^2 value for a plot against SlogP is only 0.011. (For a plot in linear space the value drops to just r^2 = 0.004, data not shown.) There is a slightly clearer relationship between Caco-2 permeability and a drug's total polar surface area, but again the relationship is fairly weak (r^2 = 0.334 when the ordinate is in log space, Fig 3C, but only r^2 = 0.137 when the ordinate is in linear space (plot not shown)). It is also of interest that there is no significant relationship between total Polar Surface Area and S logP (Fig 3D). In particular, as before, we (e.g. (Dobson & Kell 2008; Kell & Oliver 2014)) and others (e.g. (Skolnik et al. 2010)) find that transmembrane permeability cannot be accounted for in terms of simple biophysical properties, and certainly not via logP.

Lack of relationship between Caco-2 permeability and structural similarity to endogenous metabolites

Since the natural role of the transporters that drugs hitchhike on is to transport endogenous metaboliltes (Dobson & Kell 2008; Kell 2013; Kell 2015; Kell et al. 2013; Kell & Oliver 2014; Nigam 2015; Swainston et al. 2013), the 'principle of molecular similarity' (e.g. (Bender & Glen 2004; Eckert & Bajorath 2007; Gasteiger 2003; Maldonado et al. 2006)) suggests that drugs should bear structural similarities to endogenous metabolites, and this is found to be the case (Dobson et al. 2009b; O'Hagan & Kell 2015b; O'Hagan et al. 2015). This led us to wonder whether any aspects of 'metabolite-likeness' might be related to Caco-2 permeability. However, we found no simple relationship of this type, whether (as illustrated) in terms of the closest Tanimoto similarity (Fig 3A) or (for the 61 molecules for which this was true) the count of endogenites exceeding a Tanimoto similarity of 0.65 (Fig 3B). (There was a very weak positive correlation, r^2 = 0.156, with the number of endogenites exceeding a Tanimoto similarity of 0.75, for the 21 molecules that had at least one, data not shown.) One interpretation of this is that while in some cases a rather small number of transporters are typically involved in drug uptake (e.g. (Winter et al. 2014)), in many cases a considerably greater number contribute (e.g. (Kell et al. 2013; Lanthaler et al. 2011)). While well enough known in general (Mestres & Gregori-Puigjané 2009), such 'promiscuity' has become much more manifest using modern chemical biology approaches to detect protein binding directly (e.g. (Li et al. 2010; Niphakis et al. 2015)).

Finally, we wondered whether a standard machine learning approach (a random forest learner (Breiman 2001; Knight et al. 2009; O'Hagan & Kell 2015b)) might be able to predict Caco-2 permeabilities using a couple of fingerprint methods for encoding drug structures. Even this very powerful method had negligible predictive power as judged by its out-of-bag error (Fig 5). It must be concluded that the ability to pass through Caco-2 cells is a very heterogeneous property, that cannot be accounted for via simple biophysical properties (e.g. those contributing to log P), and is best explained by the intermediacy of a very heterogeneous set of transporters.

Discussion and conclusions

A recent publication (Matsson et al. 2015a; Matsson et al. 2015b), using exceptionally high values of P_{app} for verapamil and propranolol, claimed that the apparent permeability values were such that they could not be supported by known (random) transporters at random expression level, K_m and k_{cat} values. It was stated (Matsson et al. 2015a) that such rates "are possible in the absence of transmembrane diffusion, but only under very specific conditions that rarely or never occur for known human drug transporters". While we showed that this was simply not the case (quite the opposite) (Mendes et al. 2015), it prompted us to ask the question as to what typical rates of P_{app} might be for marketed drugs more generally. By bringing together tabulated data from 15 studies, we found that the commonest values are just ca 3-4 . 10^{-6} cm.s^{-1}, and that the median value is ca 16 . 10^{-6} cm.s^{-1}. Thus, transporters alone can easily account for these. There was no significant correlation of P_{app} values with either the values of various biophysical descriptors or measures of endogenite-likeness, and even powerful machine learning methods could not predict the permeabilities from the drug structures. The most obvious reason for this is simply that there is no unitary explanation (such as simplistic phospholipid bilayer diffusion), as most drugs exploit multiple but often unknown transporters with overlapping specificities. Which they are and how much each contributes to a given Caco-2 permeability must be determined by varying their activities as independent variables (Kell 2015; Kell & Oliver 2014; Kell et al. 2015), whether by using inhibitors (e.g. (Han et al. 2015; Ming et al. 2009)) or genetically. This latter activity has been initiated in other cell lines (e.g. (Giacomini et al. 2010; Han et al. 2015; Lanthaler et al. 2011; Winter et al. 2014)). The availability of powerful mammalian genome editing tools such as variants of the CRISPR/Cas9 system (e.g. (Kleinstiver et al. 2015; Maeder et al. 2013; Wang et al. 2014; Zhou et al. 2014)) imply that we may soon expect to see this strategy applied with great effect to the Caco-2 system.

Funding statement

We thank the Biotechnology and Biological Sciences Research Council (BBSRC) for financial support under grant BB/M017702/1. This is a contribution from the Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM).

Figures and legends to figures

Figure 1. A compilation of 15 review articles on Caco-2 permeability measurements. **A**. Full dataset, including outliers. **B**. Reduced dataset after removal of the data from (Avdeef et al. 2005). **C**. Cumulative plot and smoothed histogram of the Caco-2 permeabilities in the reduced dataset. In Figure 1C data for identical drugs were averaged. Data were extracted from the following papers. 1 (Bergström et al. 2003); 2 (Hou et al. 2004); 3 (Corti et al. 2006); 4 (Balimane et al. 2006); 5 (Gozalbes et al. 2011); 6 (Peng et al. 2014); 7 (Press 2011); 8 (Usansky & Sinko 2005); 9 (Marino et al. 2005); 10 (Avdeef et al. 2005); 11 (Hayeshi et al. 2008); 12 (Wang et al. 2010); 13 (Uchida et al. 2009); 14 (Skolnik et al. 2010); 15 (Lin et al. 2011)

Figure 2. Relative lack of effect of efflux inhibitors on Caco-2 permeabilities of marketed drugs. Data are taken from (Lin et al. 2011) and shown as paired values.

Figure 3. Lack of relationship between Caco-2 cells and simple biophysical parameters. **A**. Caco-2 permeability as a function of MW. **B**. Caco-2 permeability as a function of SlogP. **C**. Caco-2 permeability as a function of Total Polar Surface Area. **D**. Lack of relationship between Total Polar Surface Area and S log P.

Figure 4. Lack of relationship between Caco-2 cell permeability and measures of endogenite-likeness. **A.** Lack of relationship between the P_{app} of a drug in Caco-2 cells and its greatest Tanimoto similarity to any endogenite molecule in Recon2. **B.** Lack of relationship between the P_{app} of a drug and the number of endogenous metabolites (endogenites) in Recon2 possessing a Tanimoto similarity greater than 0.65.

Figure 5. Lack of relationship between experimental Caco-2 permeabilities and those predicted (via out-of-bag estimation) from a random forest learner. Drug properties were encoded using either the MACCS166 encoding (O'Hagan et al. 2015) or the full DES encoding (O'Hagan & Kell 2015b), each together with the molecular properties encoded in the CDK KNIME node (Beisken et al. 2013).

Supplemental Table S1.

Set of Caco-2 permeabilities and RDKit descriptors used herein. Excel file.
Drugs_Caco2_compilation_with_descriptors_2.xls

References

Anderle P, Huang Y, and Sadée W. 2004. Intestinal membrane transport of drugs and nutrients: genomics of membrane transporters using expression microarrays. *Eur J Pharm Sci* 21:17-24.

Artursson P, Palm K, and Luthman K. 2001. Caco-2 monolayers in experimental and theoretical predictions of drug transport. *Adv Drug Deliv Rev* 46:27-43.

Avdeef A. 2012. *Absorption and drug development: solubility, permeability and charge state*. New York: Wiley.

Avdeef A, Artursson P, Neuhoff S, Lazorova L, Gråsjö J, and Tavelin S. 2005. Caco-2 permeability of weakly basic drugs predicted with the double-sink PAMPA pKa(flux) method. *Eur J Pharm Sci* 24:333-349.

Awortwe C, Fasinu PS, and Rosenkranz B. 2014. Application of Caco-2 cell line in herb-drug interaction studies: current approaches and challenges. *J Pharm Pharm Sci* 17:1-19.

Balimane PV, and Chong S. 2005. Cell culture-based models for intestinal permeability: a critique. *Drug Discov Today* 10:335-343.

Balimane PV, Han YH, and Chong SH. 2006. Current industrial practices of assessing permeability and P-glycoprotein interaction. *Aaps J* 8:E1-E13.

Beisken S, Meinl T, Wiswedel B, de Figueiredo LF, Berthold M, and Steinbeck C. 2013. KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics* 14:257.

Bender A, and Glen RC. 2004. Molecular similarity: a key technique in molecular informatics. *Org Biomol Chem* 2:3204-3218.

Bergström CAS, Strafford M, Lazorova L, Avdeef A, Luthman K, and Artursson P. 2003. Absorption classification of oral drugs based on molecular surface properties. *J Med Chem* 46:558-570.

Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Sieb C, Thiel K, and Wiswedel B. 2008. KNIME: the Konstanz Information Miner. In: Preisach C, Burkhardt H, Schmidt-Thieme L, and Decker R, eds. *Data Analysis, Machine Learning and Applications*. Berlin: Springer, 319-326.

Bhardwaj RK, Herrera-Ruiz DR, Xu Y, Carl SM, Cook TJ, Vorsa N, and Knipp GT. 2008. Intestinal transporters in drug absorption. In: Krishna R, and Yu L, eds. *Biopharmaceutics Applications in Drug Development*. Berlin: Springer, 175-261.

Breiman L. 2001. Random forests. *Machine Learning* 45:5-32.

Cao X, Yu LX, and Sun D. 2006. Drug Absorption Principles. In: Krishna R, and Yu L, eds. *Biopharmaceutics Applications in Drug Development*. New York: Springer, 75-100.

Corti G, Maestrelli F, Cirri M, Zerrouk N, and Mura P. 2006. Development and evaluation of an *in vitro* method for prediction of human drug absorption - II. Demonstration of the method suitability. *Eur J Pharm Sci* 27:354-362.

Dobson P, Lanthaler K, Oliver SG, and Kell DB. 2009a. Implications of the dominant role of cellular transporters in drug uptake. *Curr Top Med Chem* 9:163-184.

Dobson PD, and Kell DB. 2008. Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat Rev Drug Disc* 7:205-220.

Dobson PD, Patel Y, and Kell DB. 2009b. "Metabolite-likeness" as a criterion in the design and selection of pharmaceutical drug libraries. *Drug Disc Today* 14:31-40.

Ecker G, and Chiba P. 2009. Transporters as drug carriers: structure, function, substrates. Weinheim: Wiley/VCH.

Eckert H, and Bajorath J. 2007. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today* 12:225-233.

Fearn RA, and Hirst BH. 2006. Predicting oral drug absorption and hepatobiliary clearance: Human intestinal and hepatic *in vitro* cell models. *Env Toxicol Pharmacol* 21:168-178.

Feng B, Varma MV, Costales C, Zhang H, and Tremaine L. 2014. *In vitro* and *in vivo* approaches to characterize transporter-mediated disposition in drug discovery. *Expert Opin Drug Discov* 9:873-890.
Fromm MF, and Kim RB. 2011. Drug Transporters. Handbook of Experimental Pharmacology Berlin: Springer.
Gasteiger J. 2003. Handbook of Chemoinformatics: From Data to Knowledge. Weinheim: Wiley/VCH.
Giacomini KM, Huang SM, Tweedie DJ, Benet LZ, Brouwer KL, Chu X, Dahlin A, Evers R, Fischer V, Hillgren KM, Hoffmaster KA, Ishikawa T, Keppler D, Kim RB, Lee CA, Niemi M, Polli JW, Sugiyama Y, Swaan PW, Ware JA, Wright SH, Wah Yee S, Zamek-Gliszczynski MJ, and Zhang L. 2010. Membrane transporters in drug development. *Nat Rev Drug Discov* 9:215-236.
Gozalbes R, Jacewicz M, Annand R, Tsaioun K, and Pineda-Lucena A. 2011. QSAR-based permeability model for drug-like compounds. *Bioorg Med Chem* 19:2615-2624.
Han TK, Proctor WR, Costales CL, Cai H, Everett RS, and Thakker DR. 2015. Four cation-selective transporters contribute to apical uptake and accumulation of metformin in Caco-2 cell monolayers. *J Pharmacol Exp Ther* 352:519-528.
Hayeshi R, Hilgendorf C, Artursson P, Augustijns P, Brodin B, Dehertogh P, Fisher K, Fossati L, Hovenkamp E, Korjamo T, Masungi C, Maubon N, Mols R, Müllertz A, Mönkkönen J, O'Driscoll C, Oppers-Tiemissen HM, Ragnarsson EG, Rooseboom M, and Ungell AL. 2008. Comparison of drug transporter gene expression and functionality in Caco-2 cells from 10 different laboratories. *Eur J Pharm Sci* 35:383-396.
Hidalgo IJ, Raub TJ, and Borchardt RT. 1989. Characterization of the human colon carcinoma cell line (Caco-2) as a model system for intestinal epithelial permeability. *Gastroenterology* 96:736-749.
Hilgendorf C, Ahlin G, Seithel A, Artursson P, Ungell AL, and Karlsson J. 2007. Expression of thirty-six drug transporter genes in human intestine, liver, kidney, and organotypic cell lines. *Drug Metab Dispos* 35:1333-1340.
Hou TJ, Zhang W, Xia K, Qiao XB, and Xu XJ. 2004. ADME evaluation in drug discovery. 5. Correlation of Caco-2 permeation with simple molecular properties. *J Chem Inf Comput Sci* 44:1585-1600.
Ishikawa T, Kim RB, and König J. 2013. Pharmacogenomics of Human Drug Transporters: Clinical Impacts New York: Wiley.
Kell DB. 2013. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening, and knowledge of transporters: where drug discovery went wrong and how to fix it. *Febs J* 280:5957-5980.
Kell DB. 2015. What would be the observable consequences if phospholipid bilayer diffusion of drugs into cells is negligible? *Trends Pharmacol Sci* 36:15-21.
Kell DB, and Dobson PD. 2009. The cellular uptake of pharmaceutical drugs is mainly carrier-mediated and is thus an issue not so much of biophysics but of systems biology. In: Hicks MG, and Kettner C, eds. *Proc Int Beilstein Symposium on Systems Chemistry*. Berlin: Logos Verlag, 149-168 - http://www.beilstein-institut.de/Bozen2008/Proceedings/Kell/Kell.pdf.
Kell DB, Dobson PD, Bilsland E, and Oliver SG. 2013. The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: what we (need to) know and how we can do so. *Drug Disc Today* 18:218-239.
Kell DB, Dobson PD, and Oliver SG. 2011. Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only. *Drug Disc Today* 16:704-714.
Kell DB, and Goodacre R. 2014. Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery. *Drug Disc Today* 19:171-182.
Kell DB, and Oliver SG. 2014. How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion. *Front Pharmacol* 5:231.

Kell DB, Swainston N, Pir P, and Oliver SG. 2015. Membrane transporter engineering in industrial biotechnology and whole-cell biocatalysis. *Trends Biotechnol* 33:237-246.

Kleinstiver BP, Prew MS, Tsai SQ, Topkar VV, Nguyen NT, Zheng Z, Gonzales APW, Li Z, Peterson RT, Yeh JR, Aryee MJ, and Joung JK. 2015. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523:481-485.

Knight CG, Platt M, Rowe W, Wedge DC, Khan F, Day P, McShea A, Knowles J, and Kell DB. 2009. Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape. *Nucleic Acids Res* 37:e6.

Krogsgaard-Larsen P, Liljefors T, and Madsen U. 1996. A textbook of drug design and development, 2nd Ed. New York: Harwod Academic Publishers.

Landowski CP, Anderle P, Sun D, Sadee W, and Amidon GL. 2004. Transporter and ion channel gene expression after Caco-2 cell differentiation using 2 different microarray technologies. *Aaps J* 6:e21.

Lanthaler K, Bilsland E, Dobson P, Moss HJ, Pir P, Kell DB, and Oliver SG. 2011. Genome-wide assessment of the carriers involved in the cellular uptake of drugs: a model system in yeast. *BMC Biol* 9:70.

Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, and Wishart DS. 2014. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* 42:D1091-1097.

Li X, Gianoulis TA, Yip KY, Gerstein M, and Snyder M. 2010. Extensive *in vivo* metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* 143:639-650.

Lin X, Skolnik S, Chen X, and Wang J. 2011. Attenuation of intestinal absorption by major efflux transporters: quantitative tools and strategies using a Caco-2 model. *Drug Metab Dispos* 39:265-274.

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, and Joung JK. 2013. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10:977-979.

Maldonado AG, Doucet JP, Petitjean M, and Fan BT. 2006. Molecular similarity and diversity in chemoinformatics: from theory to applications. *Mol Divers* 10:39-79.

Marino AM, Yarde M, Patel H, Chong S, and Balimane PV. 2005. Validation of the 96 well Caco-2 cell culture model for high throughput permeability assessment of discovery compounds. *Int J Pharm* 297:235-241.

Matsson P, Fenu LA, Lundquist P, Wisńiewski JR, Kansy M, and Artursson P. 2015a. Quantifying the impact of transporters on cellular drug permeability. *Trends Pharmacol Sci* 36:255-262.

Matsson P, Fenu LA, Lundquist P, Wisńiewski JR, Kansy M, and Artursson P. 2015b. Supplementary Information: addendum to 'Quantifying the impact of transporters on cellular drug permeability'. *Trends Pharmacol Sci* 36:in press.

Mazanetz MP, Marmon RJ, Reisser CBT, and Morao I. 2012. Drug discovery applications for KNIME: an open source data mining platform. *Curr Top Med Chem* 12:1965-1979.

Mendes P, Oliver SG, and Kell DB. 2015. Fitting transporter activities to cellular drug concentrations and fluxes: why the bumblebee can fly. *Trends Pharmacol Sci*:in press.

Mestres J, and Gregori-Puigjané E. 2009. Conciliating binding efficiency and polypharmacology. *Trends Pharmacol Sci* 30:470-474.

Ming X, Ju W, Wu H, Tidwell RR, Hall JE, and Thakker DR. 2009. Transport of dicationic drugs pentamidine and furamidine by human organic cation transporters. *Drug Metab Dispos* 37:424-430.

Nigam SK. 2015. What do drug transporters really do? *Nat Rev Drug Discov* 14:29-44.

Niphakis MJ, Lum KM, Cognetta AB, Correia BE, Ichu TA, Olucha J, Brown SJ, Kundu S, Piscitelli F, Rosen H, and Cravatt BF. 2015. A Global Map of Lipid-Binding Proteins and Their Ligandability in Cells. *Cell* 161:1668-1680.

O'Hagan S, and Kell DB. 2015a. Software review: The KNIME workflow environment and its applications in Genetic Programming and machine learning. *Genetic Progr Evol Mach*:in press.

O'Hagan S, and Kell DB. 2015b. Understanding the foundations of the structural similarities between marketed drugs and endogenous human metabolites. *Front Pharmacol* 6:105.

O'Hagan S, Swainston N, Handl J, and Kell DB. 2015. A ‘rule of 0.5’ for the metabolite-likeness of approved pharmaceutical drugs. *Metabolomics* 11:323-339.

Peng Y, Yadava P, Heikkinen AT, Parrott N, and Railkar A. 2014. Applications of a 7-day Caco-2 cell model in drug discovery and development. *Eur J Pharm Sci* 56:120-130.

Press B. 2011. Optimization of the Caco-2 permeability assay to screen drug compounds for intestinal absorption and efflux. *Methods Mol Biol* 763:139-154.

Pshezhetsky AV, Fedjaev M, Ashmarina L, Mazur A, Budman L, Sinnett D, Labuda D, Beaulieu JF, Menard D, Nifant'ev I, and Levy E. 2007. Subcellular proteomics of cell differentiation: quantitative analysis of the plasma membrane proteome of Caco-2 cells. *Proteomics* 7:2201-2215.

Riniker S, and Landrum GA. 2013. Open-source platform to benchmark fingerprints for ligand-based virtual screening. *J Cheminform* 5:26.

Rubas W, Cromwell ME, Shahrokh Z, Villagran J, Nguyen TN, Wellton M, Nguyen TH, and Mrsny RJ. 1996. Flux measurements across Caco-2 monolayers may predict transport in human large intestinal tissue. *J Pharm Sci* 85:165-169.

Sahoo S, Aurich MK, Jonsson JJ, and Thiele I. 2014. Membrane transporters in a human genome-scale metabolic knowledgebase and their implications for disease. *Frontiers in physiology* 5:91.

Sarmento B, Andrade F, da Silva SB, Rodrigues F, das Neves J, and Ferreira D. 2012. Cell-based *in vitro* models for predicting drug permeability. *Expert Opin Drug Metab Toxicol* 8:607-621.

Saubern S, Guha R, and Baell JB. 2011. KNIME Workflow to Assess PAINS Filters in SMARTS Format. Comparison of RDKit and Indigo Cheminformatics Libraries. *Mol Inform* 30:847-850.

Skolnik S, Lin X, Wang J, Chen XH, He T, and Zhang B. 2010. Towards prediction of in vivo intestinal absorption using a 96-well Caco-2 assay. *J Pharm Sci* 99:3246-3265.

Sugiyama Y, and Steffansen B. 2013. Transporters in Drug Development: Discovery, Optimization, Clinical Study and Regulation. New York: AAPS/Springer.

Sun D, Lennernäs H, Welage LS, Barnett JL, Landowski CP, Foster D, Fleisher D, Lee KD, and Amidon GL. 2002. Comparison of human duodenum and Caco-2 gene expression profiles for 12,000 gene sequences tags and correlation with permeability of 26 drugs. *Pharm Res* 19:1400-1416.

Sun H, Chow EC, Liu S, Du Y, and Pang KS. 2008. The Caco-2 cell monolayer: usefulness and limitations. *Expert Opin Drug Metab Toxicol* 4:395-411.

Swainston N, Mendes P, and Kell DB. 2013. An analysis of a ‘community-driven’ reconstruction of the human metabolic network. *Metabolomics* 9:757-764.

Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, Haraldsdottír H, Mo ML, Rolfsson O, Stobbe MD, Thorleifsson SG, Agren R, Bölling C, Bordel S, Chavali AK, Dobson P, Dunn WB, Endler L, Goryanin I, Hala D, Hucka M, Hull D, Jameson D, Jamshidi N, Jones J, Jonsson JJ, Juty N, Keating S, Nookaew I, Le Novère N, Malys N, Mazein A, Papin JA, Patel Y, Price ND, Selkov Sr. E, Sigurdsson MI, Simeonidis E, Sonnenschein N, Smallbone K, Sorokin A, Beek HV, Weichart D, Nielsen JB, Westerhoff HV, Kell DB, Mendes P, and Palsson BØ. 2013. A community-driven global reconstruction of human metabolism. *Nat Biotechnol* 31:419-425.

Uchida M, Fukazawa T, Yamazaki Y, Hashimoto H, and Miyamoto Y. 2009. A modified fast (4 day) 96-well plate Caco-2 permeability assay. *J Pharmacol Toxicol Methods* 59:39-43.

Usansky HH, and Sinko PJ. 2005. Estimating human drug oral absorption kinetics from Caco-2 permeability using an absorption-disposition model: model development and evaluation and derivation of analytical solutions for k_a and F_a. *J Pharmacol Exp Ther* 314:391-399.

van Breemen RB, and Li Y. 2005. Caco-2 cell permeability assays to measure drug absorption. *Expert Opin Drug Metab Toxicol* 1:175-185.

van De Waterbeemd H, and Testa B. 2009. Drug Bioavailability: Estimation of Solubility, Permeability, Absorption and Bioavailability. Weinheim: Wiley.

Volpe DA. 2011. Drug-permeability and transporter assays in Caco-2 and MDCK cell lines. *Future Med Chem* 3:2063-2077.

Wang T, Wei JJ, Sabatini DM, and Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343:80-84.

Wang Y, Cao J, Wang X, and Zeng S. 2010. Stereoselective transport and uptake of propranolol across human intestinal Caco-2 cell monolayers. *Chirality* 22:361-368.

Winter GE, Radic B, Mayor-Ruiz C, Blomen VA, Trefzer C, Kandasamy RK, Huber KVM, Gridling M, Chen D, Klampfl T, Kralovics R, Kubicek S, Fernandez-Capetillo O, Brummelkamp TR, and Superti-Furga G. 2014. The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity. *Nat Chem Biol* 10:768-773.

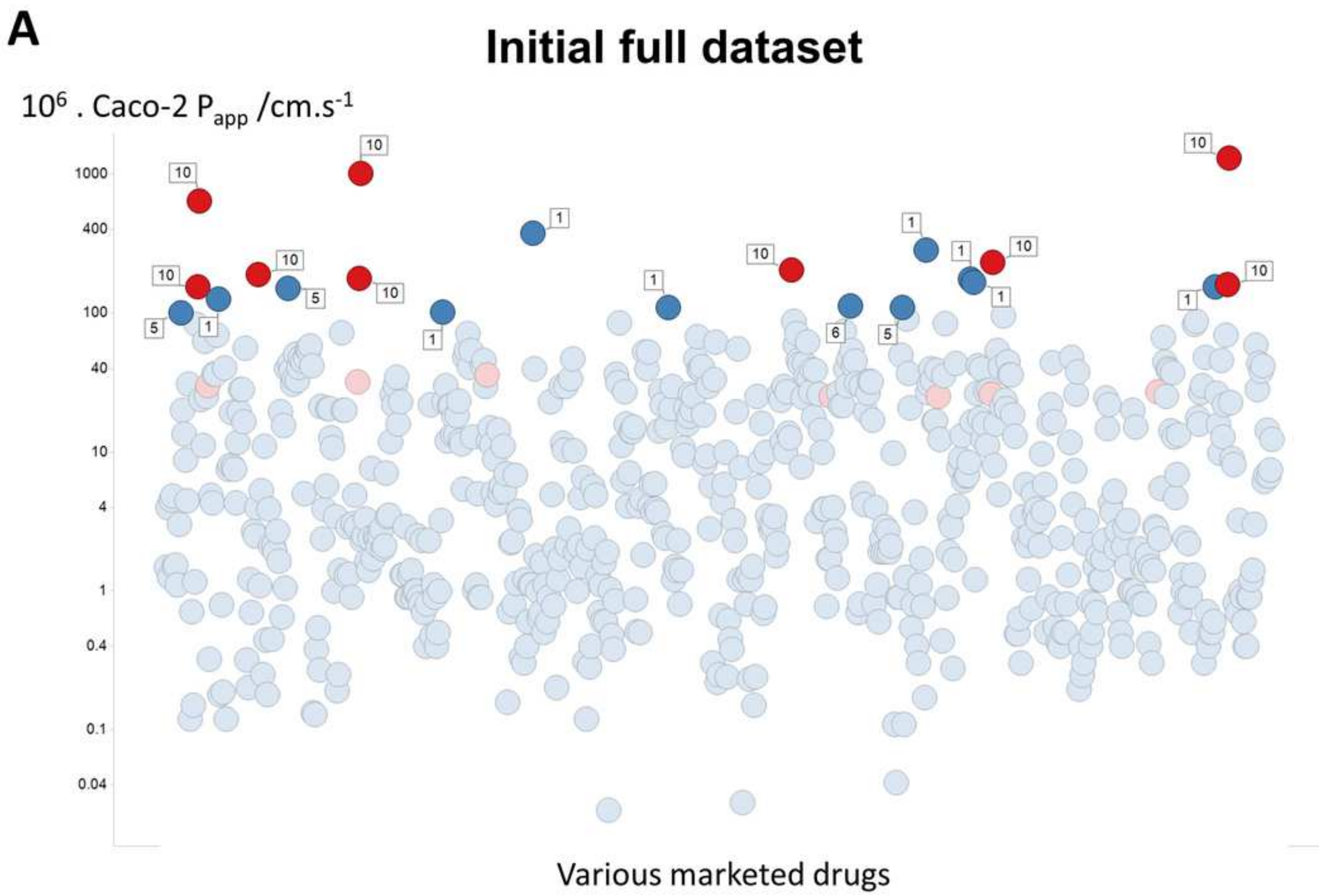
You G, and Morris ME. 2014. Drug Transporters: Molecular Characterization and Role in Drug Disposition. In: Wang B, editor. 2nd ed. New York: Wiley.

Zhou Y, Zhu S, Cai C, Yuan P, Li C, Huang Y, and Wei W. 2014. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* 509:487-491.

1

A compilation of 15 review articles on Caco-2 permeability measurements.

A. Full dataset, including outliers.

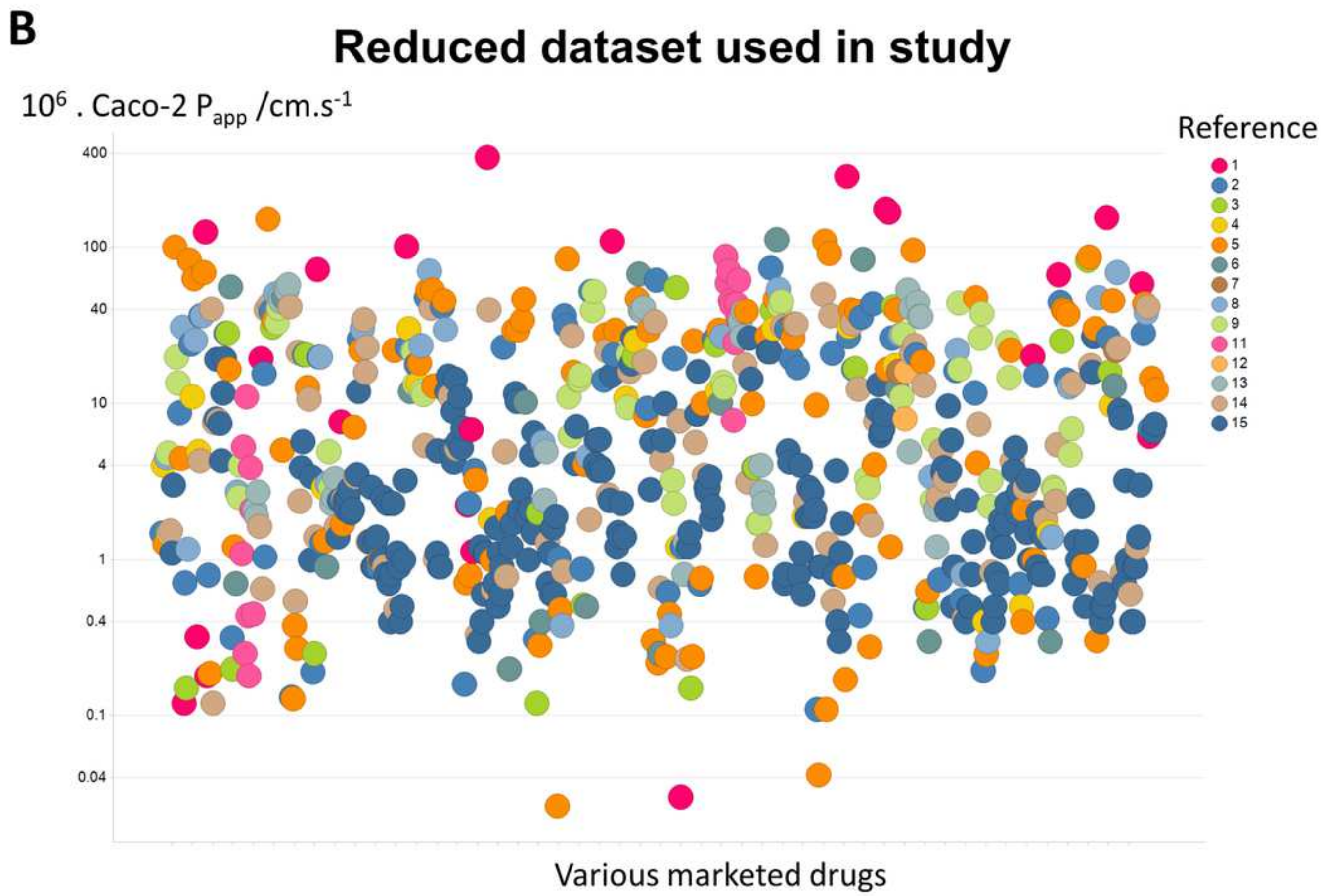


PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1271v1 | CC-BY 4.0 Open Access | rec: 29 Jul 2015, publ: 29 Jul 2015

2

A compilation of 15 review articles on Caco-2 permeability measurements.

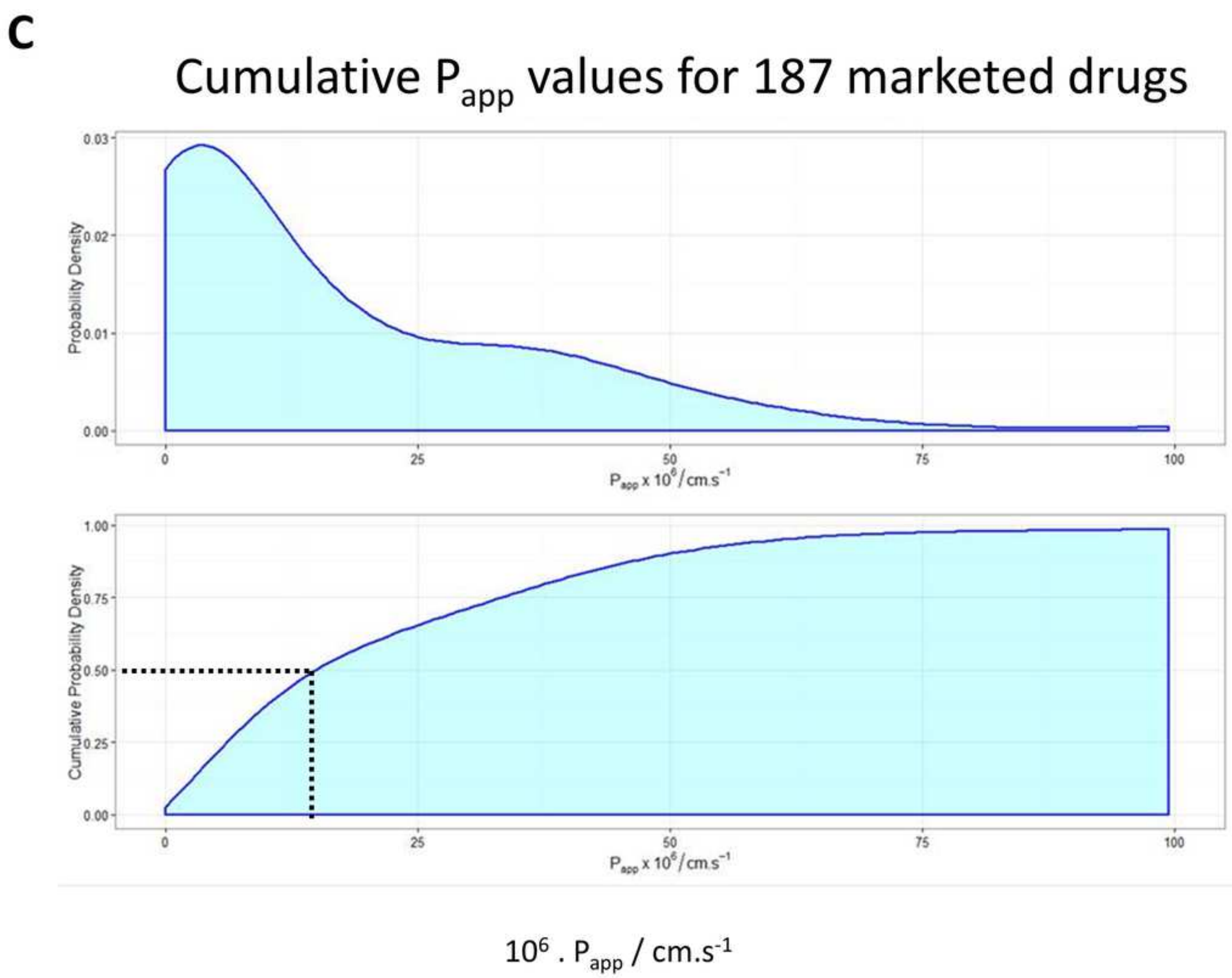
B. Reduced dataset after removal of the data from (Avdeef et al. 2005) .



3

A compilation of 15 review articles on Caco-2 permeability measurements.

C. Cumulative plot and smoothed histogram of the Caco-2 permeabilities in the reduced dataset. In Figure 1C data for identical drugs were averaged. Data were extracted from the following papers. 1 (Bergström et al. 2003) ; 2 (Hou et al. 2004) ; 3 (Corti et al. 2006) ; 4 (Balimane et al. 2006) ; 5 (Gozalbes et al. 2011) ; 6 (Peng et al. 2014) ; 7 (Press 2011) ; 8 (Usansky & Sinko 2005) ; 9 (Marino et al. 2005) ; 10 (Avdeef et al. 2005) ; 11 (Hayeshi et al. 2008) ; 12 (Wang et al. 2010) ; 13 (Uchida et al. 2009) ; 14 (Skolnik et al. 2010) ; 15 (Lin et al. 2011)

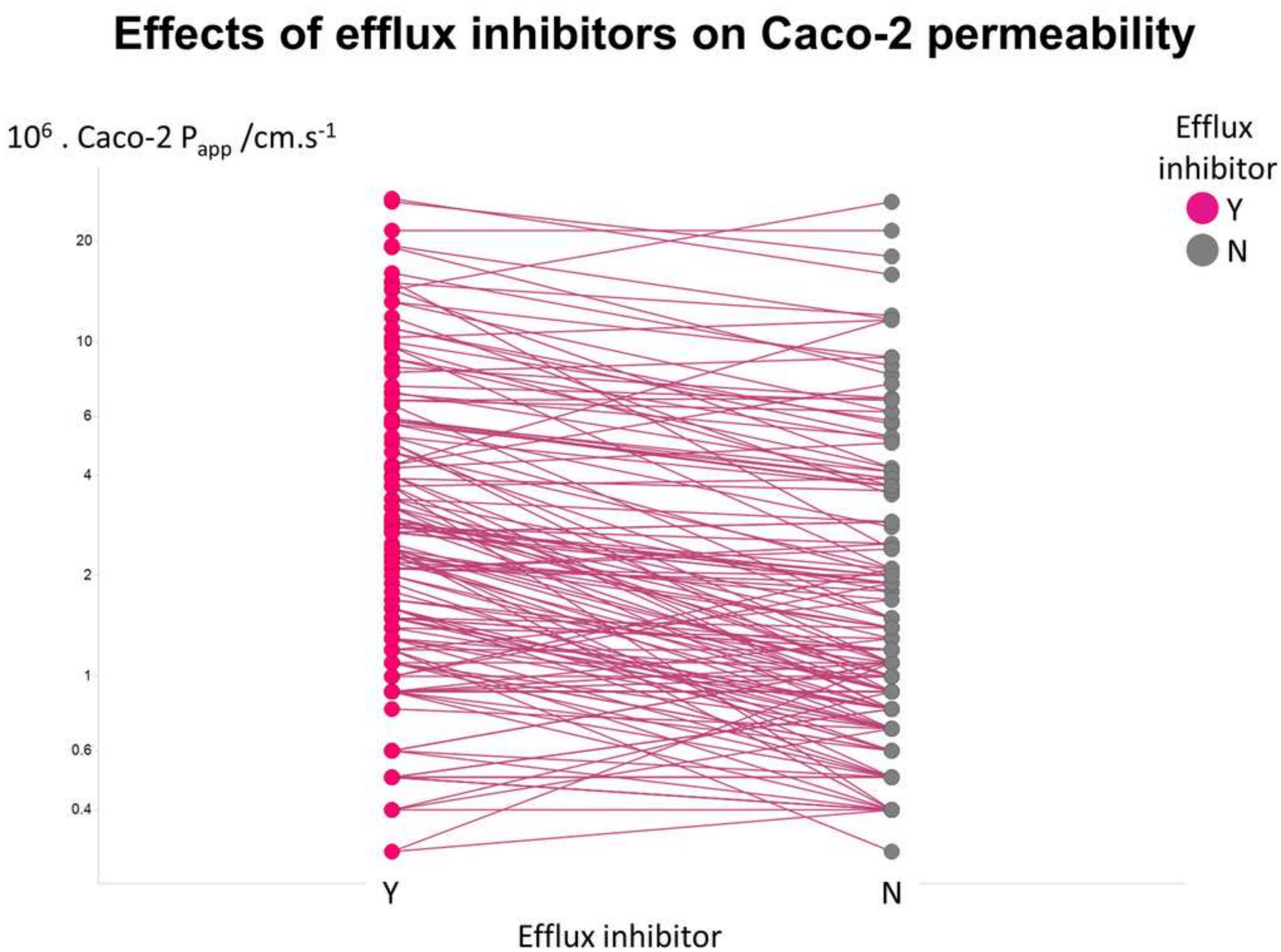


PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1271v1 | CC-BY 4.0 Open Access | rec: 29 Jul 2015, publ: 29 Jul 2015

4

Relative lack of effect of efflux inhibitors on Caco-2 permeabilities of marketed drugs.

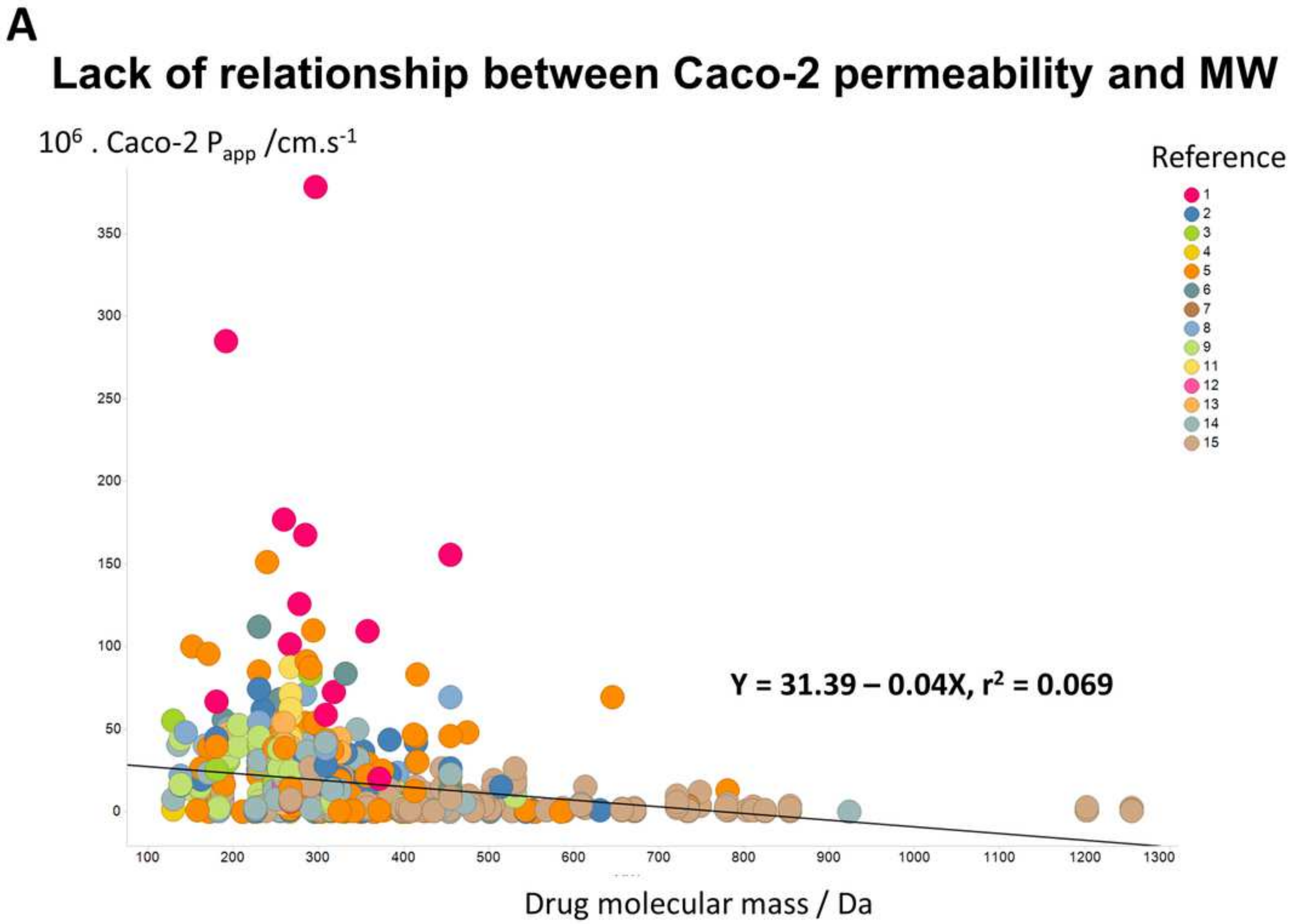
Data are taken from (Lin et al. 2011) and shown as paired values.



5

Lack of relationship between Caco-2 cells and simple biophysical parameters.

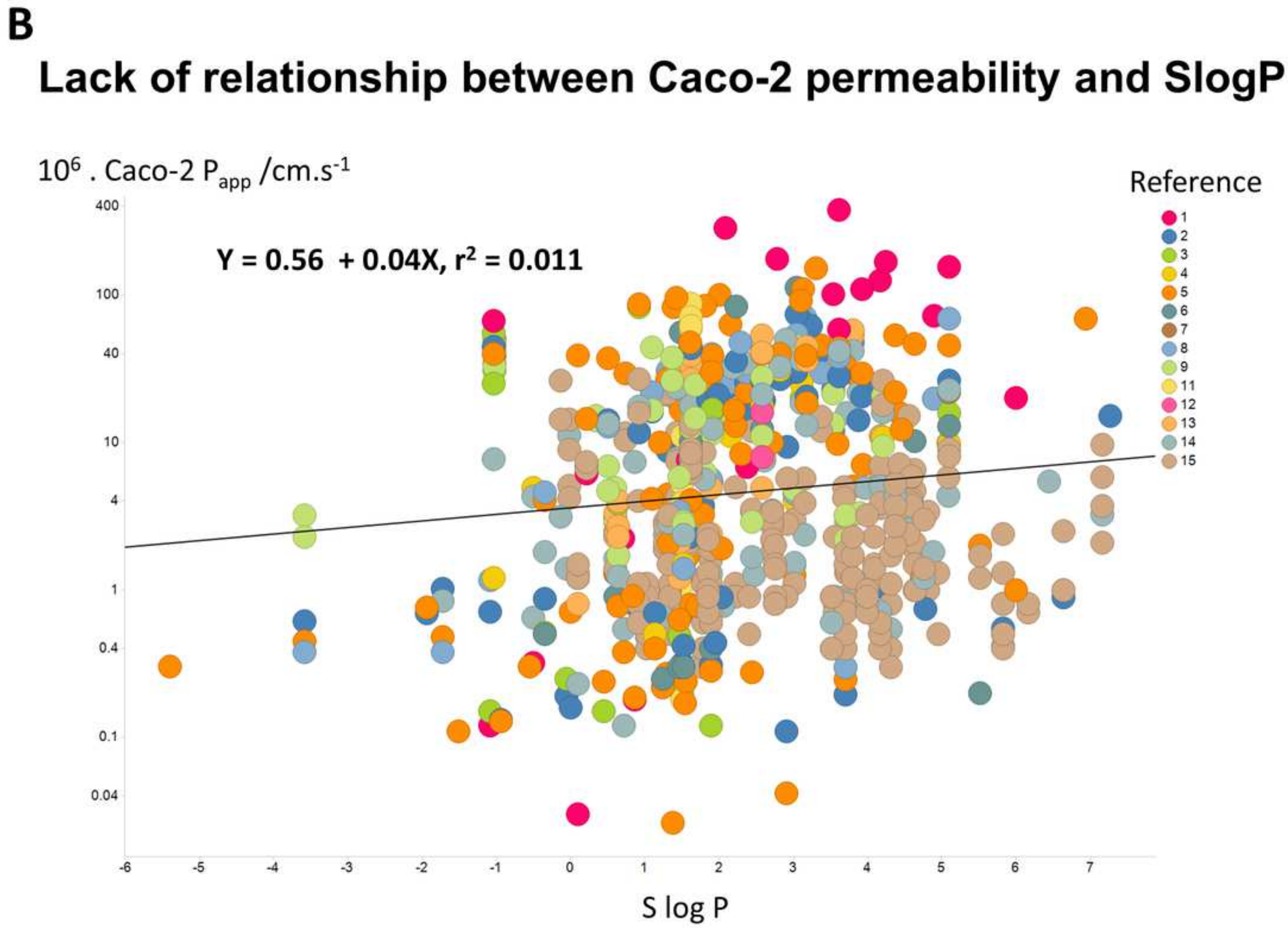
A. Caco-2 permeability as a function of MW.



6

Lack of relationship between Caco-2 cells and simple biophysical parameters .

B. Caco-2 permeability as a function of SlogP.

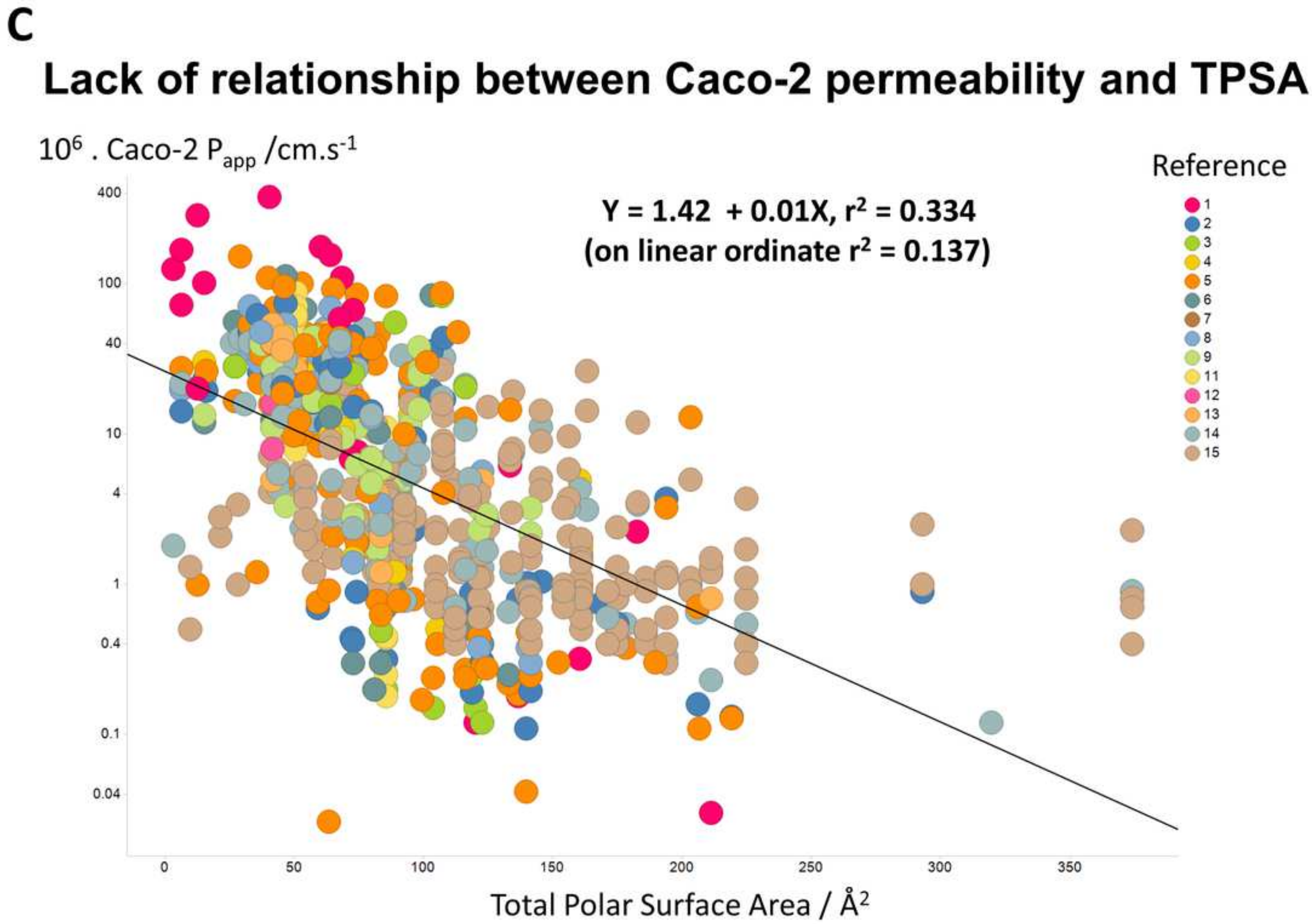


PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1271v1 | CC-BY 4.0 Open Access | rec: 29 Jul 2015, publ: 29 Jul 2015

7

Lack of relationship between Caco-2 cells and simple biophysical parameters.

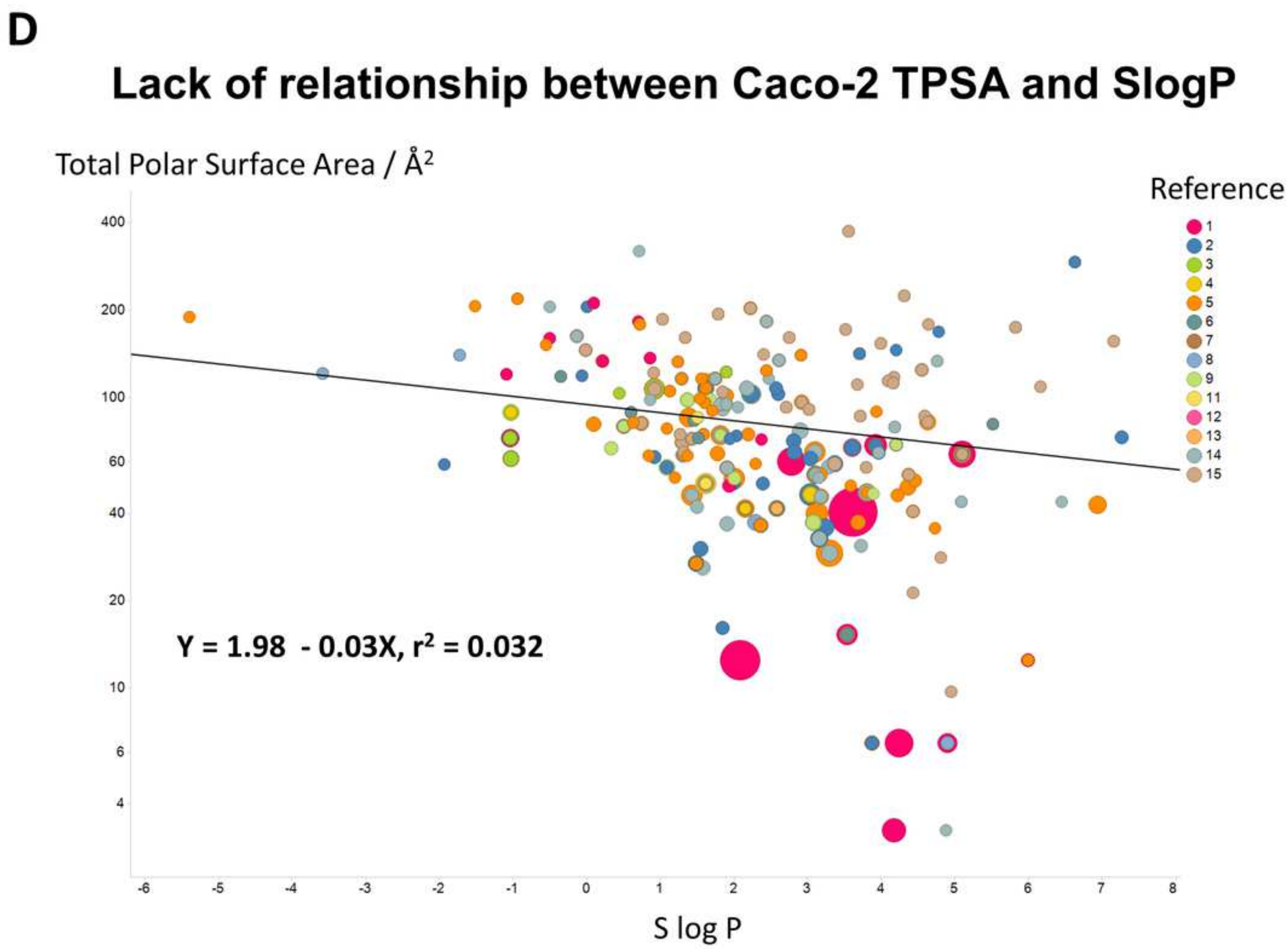
Caco-2 permeability as a function of Total Polar Surface Area.



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1271v1 | CC-BY 4.0 Open Access | rec: 29 Jul 2015, publ: 29 Jul 2015

Lack of relationship between Caco-2 cells and simple biophysical parameters.

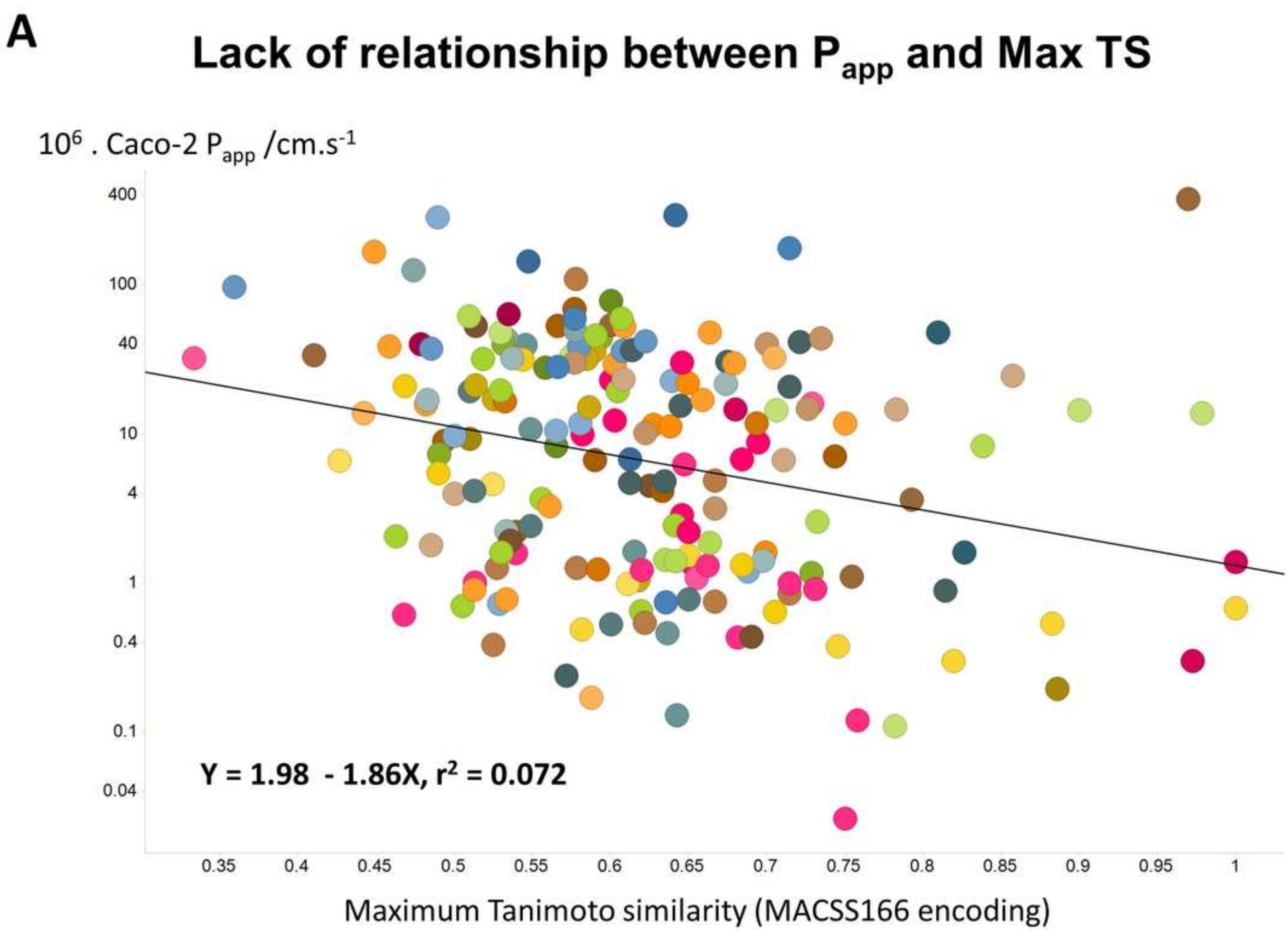
D. Lack of relationship between Total Polar Surface Area and S log P.



9

Lack of relationship between Caco-2 cell permeability and measures of endogenite-likeness.

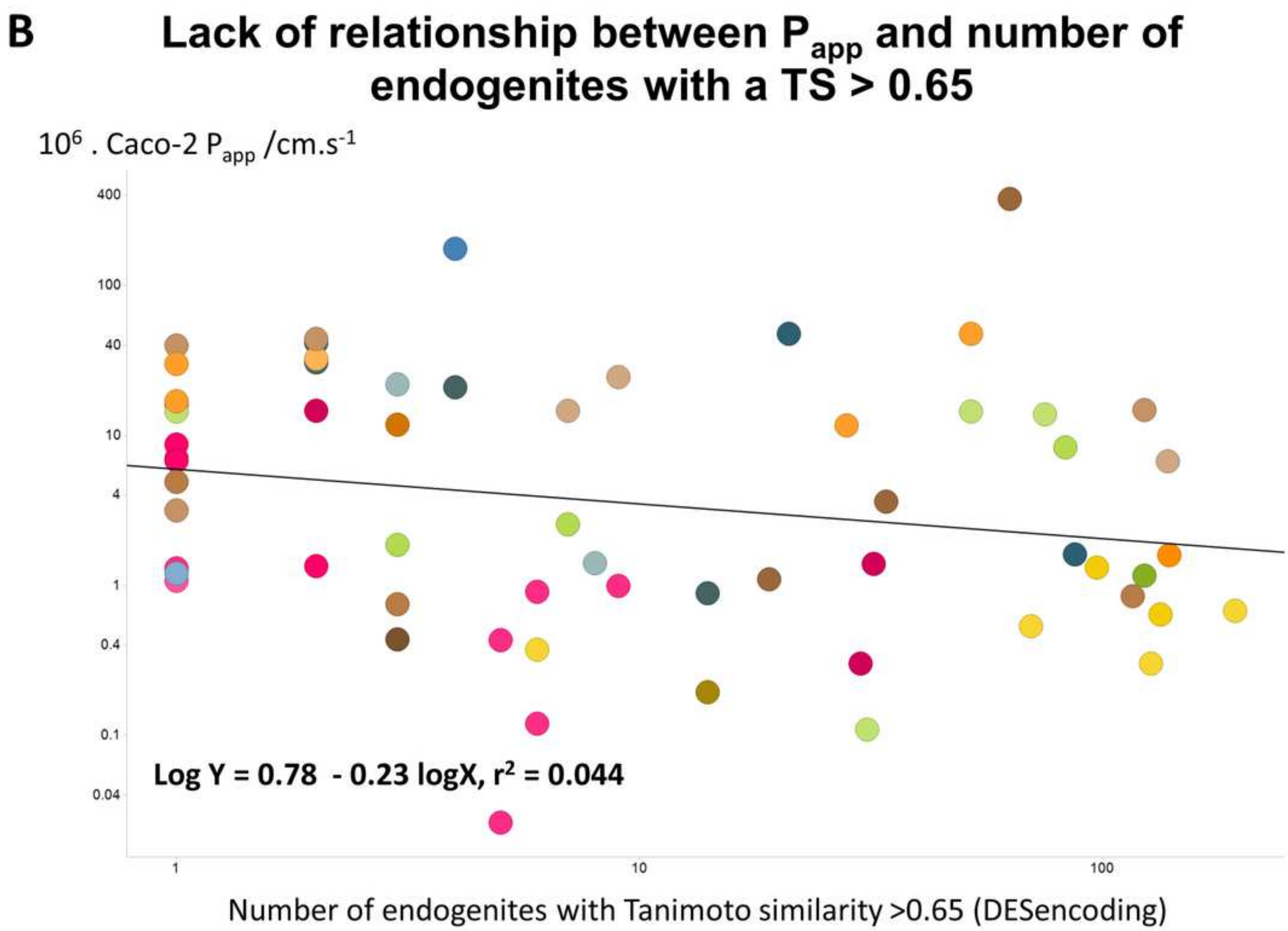
A. Lack of relationship between the P_{app} of a drug in Caco-2 cells and its greatest Tanimoto similarity to any endogenite molecule in Recon2.



10

Lack of relationship between Caco-2 cell permeability and measures of endogenite-likeness.

B. Lack of relationship between the P_{app} of a drug and the number of endogenous metabolites (endogenites) in Recon2 possessing a Tanimoto similarity greater than 0.65.



11

Lack of relationship between experimental Caco-2 permeabilities and those predicted (via out-of-bag estimation) from a random forest learner.

Drug properties were encoded using either the MACCS166 encoding (O'Hagan et al. 2015) or the full DES encoding (O'Hagan & Kell 2015b) , each together with the molecular properties encoded in the CDK KNIME node (Beisken et al. 2013) .