Review

The role and impact of high throughput biomimetic measurements in drug discovery

Shenaz Bunally* and Robert J. Young*

Drug Design & Selection, GlaxoSmithKline Medicines Research Centre, Stevenage, SG1 2NY, UK.

*Corresponding Authors E-mail: shenaz.b.bunally@gsk.com, rob.j.young@gsk.com; Tel.: +44 1738 551372

Abstract

During the early phase of drug discovery, it is becoming increasingly important to acquire the full physicochemical profile of molecules. For this purpose, there is a strong interest in developing efficient and cost-effective platforms for fast and reliable measurements of physicochemical properties. We have developed an automated physchem platform which ensures that consistent, comprehensive, and high-quality physicochemical property measurements and derived property information for 100's of compounds per week are available alongside potency data at the right time to guide compound progression decisions. We discuss the routine assessments of biomimetic properties using high throughput automated high-performance liquid chromatography (HPLC) platforms, with details of the methods and hardware employed, also with illustrations of the quality and impact of the data generated.

Keywords

Lipophilicity; Physicochemical properties; Biomimetic Chromatography, Drug Efficiency; Candidate Quality

Introduction

The use of biomimetic/physicochemical measurements, such as lipophilicity and protein/artificial membrane binding, to help rationalise the behaviour of experimental molecules in biological environments is an important facet of modern drug discovery [1,2]. Such measurements can be used not only as surrogates to model and predict behaviour, but also to estimate the quality of a given molecule and thus its chances of progression [2,3]; indeed, high clinical attrition rates have been attributed to sub-optimal physicochemical properties [4,5]. High quality, high throughput methods for biomimetic measurements are key elements of these approaches. The partitioning and distribution of drug molecules between bio-phases are fundamental to drug action [6], modelling and understanding these processes provides insight into Absorption, Distribution, Metabolism and Excretion, ADME [6], the key elements of Pharmacokinetics, the science of what the body does to a drug. The Partition and Distribution coefficients of drug molecules between 1-Octanol and aqueous buffers (OW) are well-established standards, and these biomimetic estimates of lipophilicity/hydrophobicity demonstrably influence ADME profiles and other outcomes. The negative impact of excessive lipophilicity on the chances of progression of experimental molecules has come under particular scrutiny over the past decade and changed practices in drug discovery are evident by recent improvements in the physicochemical quality of molecules [7]. Lipophilicity measurements (such as $\log_{10}$ [OW-Partition], $\log P$ or the distribution at a given pH, $\log D_{pH}$) are demonstrably unreliable for poorly

soluble compounds [8]. However, fast gradient reversed phase high pressure liquid chromatographic (HPLC) methodologies using C-18 columns provide an effective and reliable replacement [9], irrespective of solubility [10]. Analyses of data generated in this manner gave an improved resolution of ADME outcomes; together with enhanced log *P*/log *D* predictions that have enabled the construction of better structure-based in silico predictive models. Other non-silica based polymeric stationary phases are being investigated to provide insight into non-polar environments and the potential for intramolecular hydrogen bonding [11].

The grafting of stationary phases other than C-18, onto HPLC columns, enables the high throughput assessment of additional pertinent ADME interactions [12], in particular the plasma proteins human serum albumin (HSA) and alpha-1-acid-glycoprotein (AGP), plus phosphatidylcholine, which acts as a surrogate immobilised artificial membrane (IAM). The availability of these various biomimetic columns within fully automated HPLC platforms enables the high throughput and cost-effective gathering of sets of pertinent and reliable data, which can provide valuable insight on the likely behaviours in biological systems. Progressing compounds with good physicochemical properties are fundamental to pharmaceutical companies' aspirations for the objective assessment of the qualities of lead and candidate molecules [13]. An automated platform ensures that consistent, comprehensive, and high-quality physicochemical property measurements and derived property information are available at the right time to guide compound progression decisions.

The role and impact of biomimetic methods

High throughput workflow for HPLC based assays at GSK

The physicochemical/biomimetic assays are bundled to provide kinetic solubility [8], lipophilicity and biomimetic binding data on the majority of project compounds during lead discovery and optimisation. The process for preparing sample plates which are "ready to run" on the HPLC systems is shown in Figure 1. At GSK all experimental compounds are routinely dissolved and stored at 10 mM in dimethyl sulphoxide (DMSO) solution, for ease of automated handling. 50 µL samples of this stock are dispensed in 96 well master plates for the kinetic solubility assay, merged to give fully populated plates. 5 µL of the solutions in the master plates are dispensed into daughter plates; standards and blanks are added to these plates and all the samples are diluted using the appropriate solvents such as DMSO and iso-propanol/water (50/50 v/v) to produce daughter plates for the lipophilicity and biomimetic binding assays respectively. Each daughter plate has a unique barcode used to generate the plate map.

To ensure production of data of high quality and integrity, system and assay suitability checks are embedded in the sample process and data analysis. Calibration data is monitored before and after running the samples. A set of "check" standards with known lipophilicity and biomimetic binding data are run during each sequence of samples and the data checked to ensure they meet the defined criteria.

An in-house developed application (Figure 2) retrieves the relevant compound information from the barcodes and creates a "worklist" which is uploaded into the HPLC systems and used for running the samples. The application extracts the pertinent data from the raw data and places it into an Excel spreadsheet where further data analysis is conducted. The automatic extraction of data is based on a set of user-defined rules which interrogate the chromatograms and flags anomalous data, such as multiple peaks. This allows more robust and efficient data processing and analysis.

The generation of such volume of data has enabled the building of high quality in-house predictive models (discussed later in this publication) which are used for quality control. Comparison of measured data with predicted data using these models is routinely performed to highlight any anomalous data.
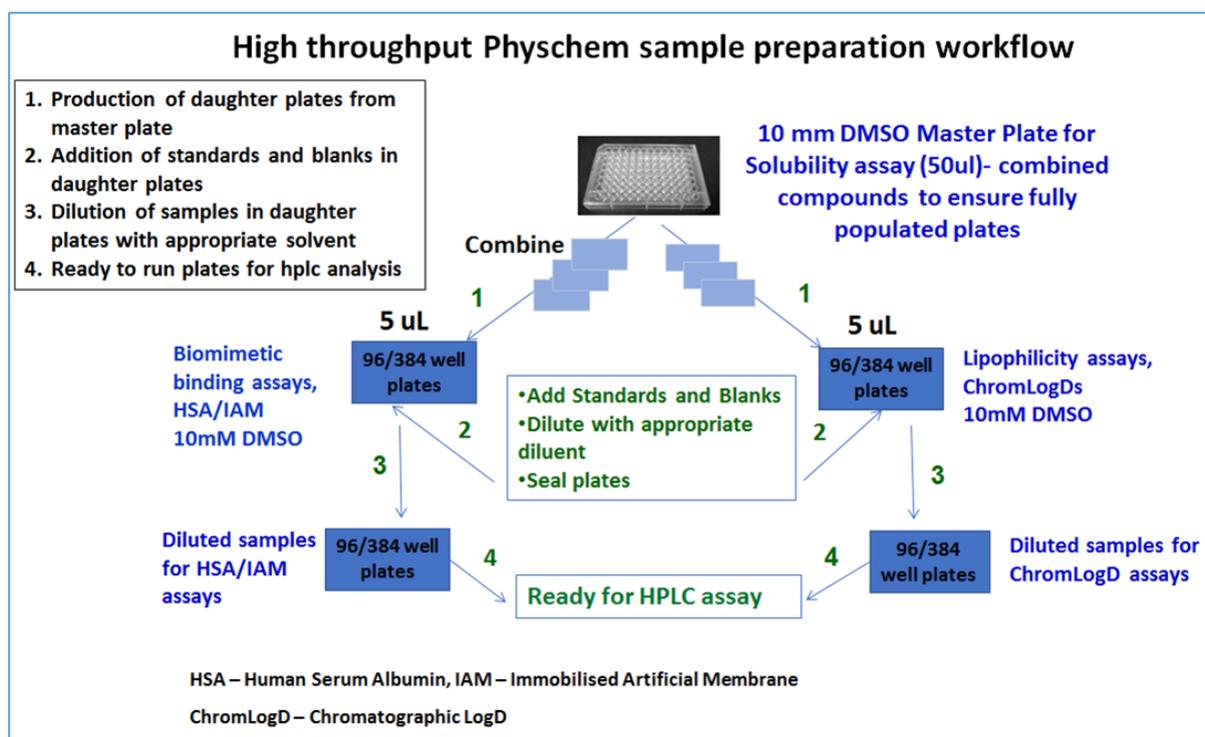
Figure 1. The GSK high throughput physchem sample preparation workflow

In addition to the kinetic solubility data, this process enables the generation of high-quality lipophilicity and biomimetic binding data for 100's of compounds to be generated weekly.
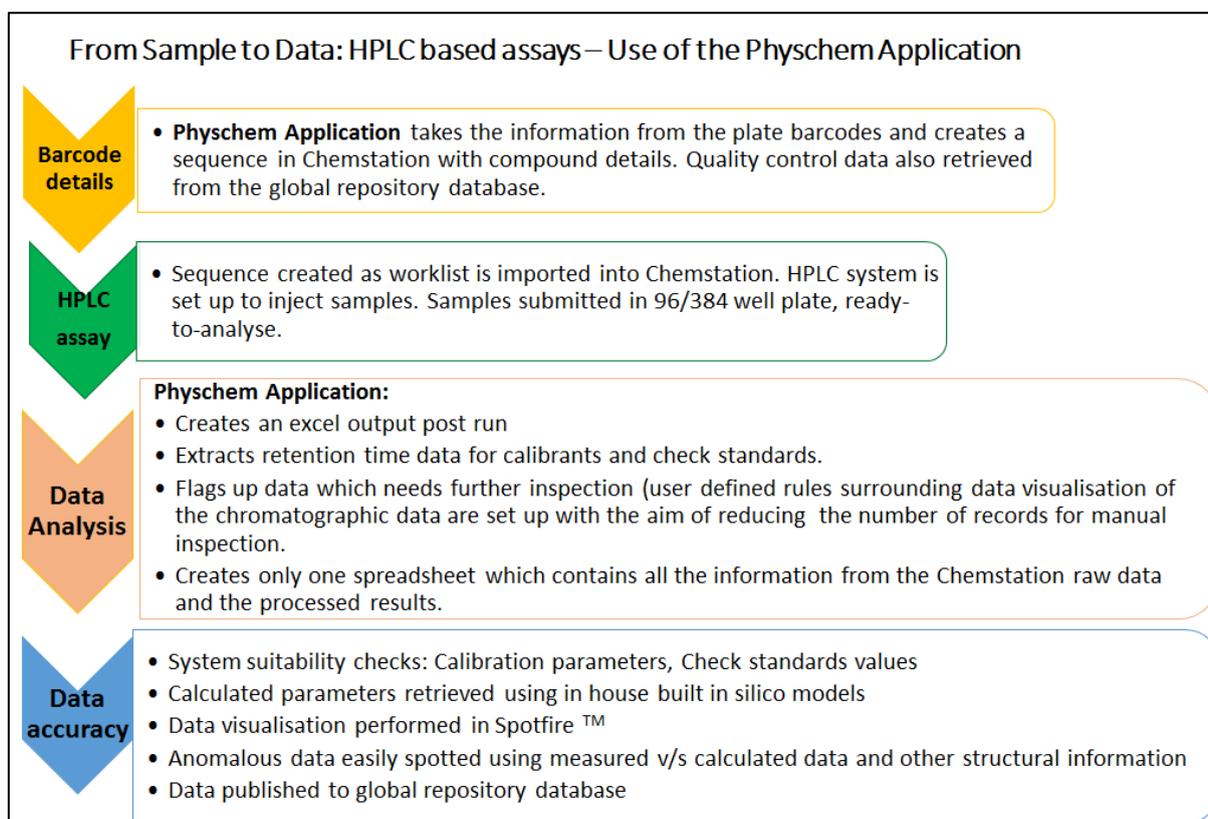
High-performance liquid chromatographic (HPLC) based assays

Horváth *et al* was amongst the first who used HPLC data for hydrophobicity measurements of amino acids [14]. It has been well described that chromatographic retention is related to the compound's dynamic distribution between the stationary and mobile phases and this is governed by a compound's hydrophobicity [15-17]. Hence HPLC offers an excellent automated platform to determine distribution coefficients of biologically active compounds between aqueous mobile phases and various non-polar and biomimetic stationary phases through measurements of retention times.

Chromatographic hydrophobicity index from fast gradient C-18 HPLC: setting new standards in lipophilicity/hydrophobicity determinations

The lipophilicity values of virtually all new compounds at GSK is measured by reversed phase HPLC using a C-18 column (50 x 2 mm 3 µM Gemini NX C18, Phenomenex, UK), at each of pH 2, 7.4 and 10.5, using buffered fast gradient acetonitrile-water mobile phases. The retention-time derived chromatographic hydrophobicity index (CHI) values are derived directly from the gradient retention times by using a calibration line obtained for standard compounds [9]. Translation of CHI values into Chrom log *D* values at the given pH is achieved using empirically-derived Equation 1 [10]. There is a deliberate offset on the scale to differentiate the data from the traditional octanol-water measurements, but there is a high correlation between the two (for soluble compounds). Figure 3 indicates how the charge profile of each compound can be estimated based on the changes in logD across the 3 pH values. For neutral molecules, the 3 are the same (i.e. the partition coefficient, Chrom log *P*); additionally, the highest Distribution constant value (for non-zwitterionic compounds) is usually a reliable estimate of the Chrom log *P* of the molecule.

Chrom log *D* = (0.0857)*CHI - 2.00 .   (1)

From Sample to Data: HPLC based assays – Use of the Physchem Application

Barcode details
- **Physchem Application** takes the information from the plate barcodes and creates a sequence in Chemstation with compound details. Quality control data also retrieved from the global repository database.

HPLC assay
- Sequence created as worklist is imported into Chemstation. HPLC system is set up to inject samples. Samples submitted in 96/384 well plate, ready-to-analyse.

Data Analysis
Physchem Application:
- Creates an excel output post run
- Extracts retention time data for calibrants and check standards.
- Flags up data which needs further inspection (user defined rules surrounding data visualisation of the chromatographic data are set up with the aim of reducing the number of records for manual inspection.
- Creates only one spreadsheet which contains all the information from the Chemstation raw data and the processed results.

Data accuracy
- System suitability checks: Calibration parameters, Check standards values
- Calculated parameters retrieved using in house built in silico models
- Data visualisation performed in Spotfire ™
- Anomalous data easily spotted using measured v/s calculated data and other structural information
- Data published to global repository database

Figure 2. Use of the physchem application for data analysis

| Chrom log D at pH:	Charge characteristic				
	Neutral	Weak acid	Strong acid	Weak base	Strong base
2	= X	= X	>X	<X	= X
7.4	X	X	X	X	X
10.5	= X	<X	= X	= X	>X

Figure 3. Patterns of changes in log D_{pH}, at high or low pH, compared to the measured value, **X**, at pH 7.4, used to classify the strength of charged motifs in compounds.

The influence and impact of these chromatographic measurements [1,10] reflect other observations on the crucial role of modulating lipophilicity in drug discovery [2,18], both through their impact in building rational understandings of both ADMET outcomes and chances of successful compound progression. Increasingly, appreciation of the impact of maximising lipophilic ligand efficiency is driving drug discovery thinking; this embodies the "Minimum lipophilicity principle" proposed by Hansch [19], who proposed that "compounds should be made as hydrophilic as possible without loss of efficacy" by subtracting lipophilicity from potency (usually expressed as ligand lipophilicity efficiency, LLE = pIC_{50} – log P) [20]. Furthermore, the principle of concurrently minimising lipophilicity and aromaticity [21] is represented by the property forecast index (PFI), the summation of aromatic ring count [22,23] and a lipophilicity measure. The measured PFI (Chrom Log $D_{7.4}$ + #Ar) is an integral part of GSK candidate quality aspirations [13], based on analyses of marketed drugs and internal attrition; ideally an oral candidate should have PFI <6, fasted-state simulated intestinal fluid (FaSSIF) solubility > 100 µg/ml and a predicted dose of < 100 mg.

Protein binding assay

Chemically bonded HSA HPLC stationary phase with column dimension of 50 x 3 mm (Chiral Technologies, France) are used for measuring compounds' binding to plasma proteins by applying linear gradient elution up to 30 % iso-propanol with 50 mM ammonium acetate buffer, pH 7.4 [24]. The gradient retention times are standardised using a calibration set of mixtures. The %HSA bound gives a reliable indication of the free fraction of the compound in plasma when compared to more complex pharmacokinetic methods. The %HSA is converted to the affinity constant, log K_{HSA}, using Equation 2:

$$\log K_{HSA} = \log [HSA\% / (101-HSA\%)] \tag{2}$$
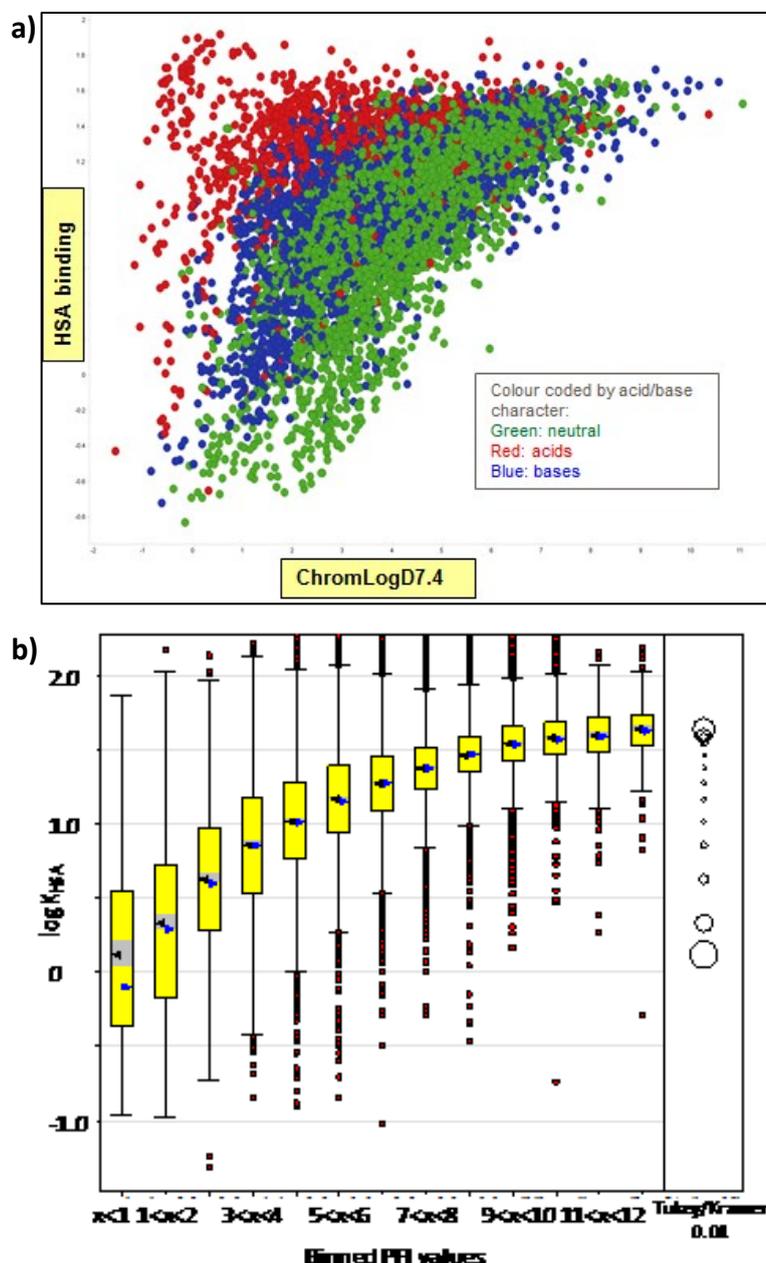
Figure 4. a) Levels of HSA binding by HPLC measurement, log K_{HSA} plotted versus Chrom log $D_{7.4}$ with charges highlighted (See Figure 3) and **b)** Box-whisker plot of log K_{HSA} vs binned measured PFI for GSK compounds, wherein K_{HSA} = [%HSA binding/(101-%HSA binding)] using the %binding values derived from chromatographic measurements.

Analysis of data derived from the HSA measurements show a clear increase in HSA binding with increasing lipophilicity (Figure 4a); when separated by charge, the increased propensity for binding by acidic

compounds, over and above their lipophilicity, is evident. The impact of aromaticity on HSA binding is also clear, given higher binding as PFI increases (Figure 4b).

Phospholipid binding assay

The binding of compounds to the immobilised artificial membrane (IAM) [25] is measured using commercially available immobilised phosphatidylcholine (PC DD2 100 x 4.6mm 10 μM, Regis Analytical, West Lafayette, USA) HPLC columns [26]. Gradient retention times obtained by applying an acetonitrile gradient up to 85 % are converted to chromatographic hydrophobicity indices (CHI$_{IAM}$) using a calibration set of compounds. The CHI$_{IAM}$ values are converted to the logarithmic retention factors using the following formula: log K_{IAM} = 0.046*CHI$_{IAM}$ + 0.42. CHI$_{IAM}$ binding gives an indication of the compound's likely binding to tissues and further insights are emerging [27], notably semi-quantitative indicators of the risk of phospholipidosis [28], a cytotoxicity outcome characterised by the breakdown of phospholipids [29], due to cationic amphiphilic drugs (CADs). The GSK model is based on the equation CAD likeness = CHI$_{IAM}$ + Delta CHI, whereby Delta CHI = (CHI$_{pH10.5}$ −CHI$_{pH7.4}$) as measured in the C-18 assays at the given pH values. Unsurprisingly, given the hydrophobic chains of the phosphatidylcholine, IAM binding is driven by lipophilicity, but, in contrast to acid-binding HSA, the net negative charge of the phosphates leads to enhanced binding of basic molecules (Figure 5).
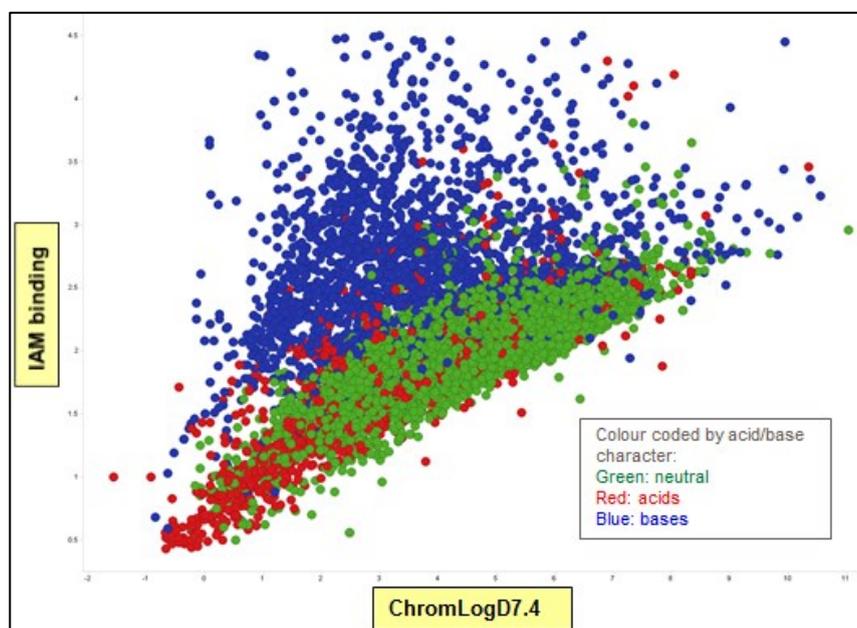
Figure 5. Levels of IAM binding by HPLC measurement (expressed as log K_{IAM} = 0.046*CHI$_{IAM}$ + 0.42) versus Chrom log $D_{7.4}$ with bases highlighted

Drug efficiency, HPLC DE$_{max}$

Drug efficiency (DE, Equation 3) is a concept based on measured pharmacokinetic parameters, designed to guide lead optimisation and developability assessment, which reflects the free plasma concentration at the site of action (expressed as the fraction of the dose) [30]. Measured DE data correlate with HPLC DE$_{max}$ values (Equation 4), generated biomimetically [31], using a combination of HSA and IAM columns Figure 6. The empirically-derived model generated from the data [32] is based on the notion that the unbound concentration is influenced by both plasma protein binding (HSA data) and the volume of distribution, for which the HPLC IAM data provides an excellent surrogate for the contributory tissue binding [33]. Increasingly, these measurements are having an impact in decision making, through estimation of clinical dose [34], and are being generated prospectively with in silico models available at GSK. The influence of

HPLC DE_{max} data on the selection of compounds for progression is illustrated in Figure 7 for a GSK programme, where most compounds had similar potencies (pIC_{50} 7 to 8) but a range of HPLC DE_{max} values. The candidates selected from this set have HPLC $DE_{max} > 1$ % and are in the same space as the profiled drugs. These HPLC measurements, pertinent to DE_{max}, are increasingly being gathered in programmes at GSK and are starting to influence thinking, design and decision making.

$$\text{DRUG Eff \%} = \frac{\text{BiophaseConc (µg/mL)}}{\text{Dose (µg/g)}} \times 100 \qquad (3)$$

$$\log \text{HPLC } DE_{max} = 2 - (0.23 * \log K_{HSA} + 0.43 * \log K_{IAM} - 0.72) \qquad (4)$$

Drug Efficiency max (DE_{max}) is the maximum in vivo drug efficiency that could theoretically be achieved assuming 100 % oral absorption, no clearance, free permeability, and no active transport.

High Potency plus High Drug Efficiency = Lower Dose Lower dose leads to reduced off-target risks. This contributes to reduced attrition.
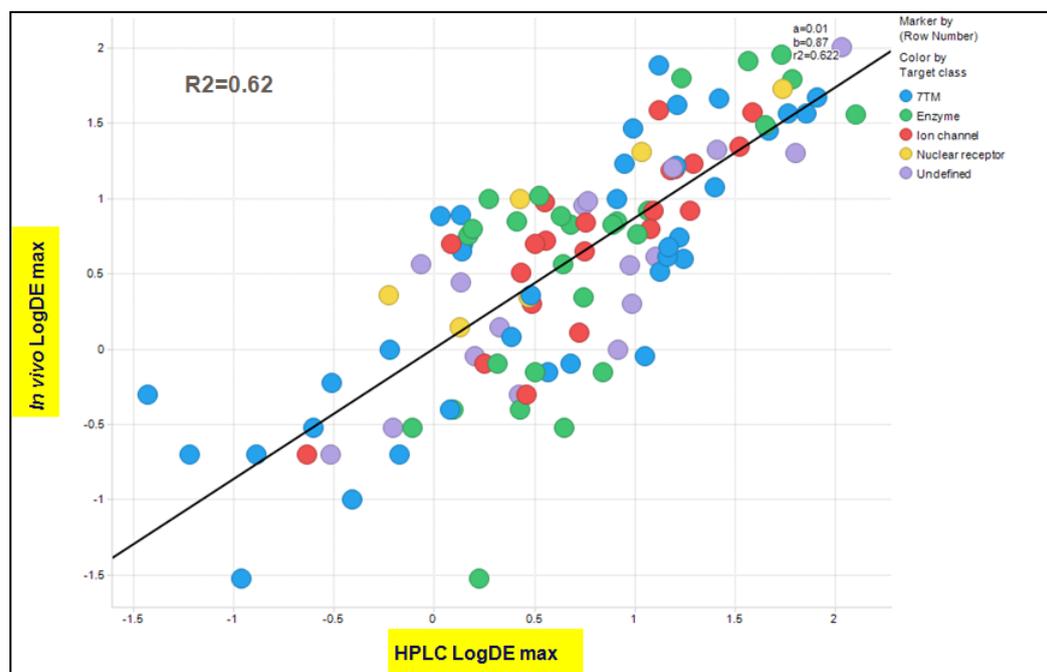

Figure 6. The plot of log (in vivo DE) vs log (HPLC DE_{max}) values for the training set of known drugs

An additional parameter, the d*rug efficiency index (DEI)* [29], can be generated by the summation of pXC_{50} and $\log_{10}$ (HPLC DE_{max}); DEI gives an estimate of the likely effective activity at the site of action, i.e. potency corrected for the free concentration.

Impact of measurements to enable, validate and improve predictive methods

Data collected by the various biomimetic measurements has enabled the building of high quality in-house predictive models of each descriptor (e.g. Figure 8 for Chrom log $D_{7.4}$). Good practice exploits iterative prediction/measurement cycles to build confidence in each series under optimisation; this also enables refinement of models on an ad hoc local basis in the rare cases that the global model does not perform well for a given structural series. The next level is to use these predictions as part of multivariate and other predictive models of various DMPK parameters (including drug efficiency), which are demonstrably improved by the enhanced predictions of physicochemical descriptors. Together, the output

of these initiatives is enabling an aspiration to *predict by first intent* in the physicochemical design process with demonstrable impacts.
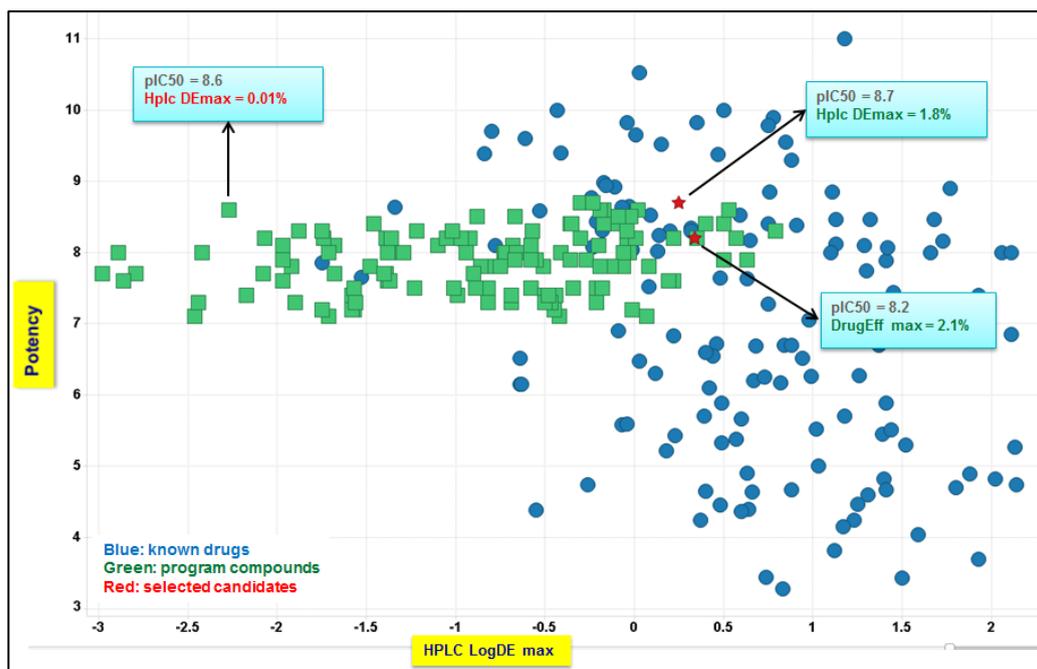
Figure 7. Plot of potency vs log (HPLC DE$_{max}$) values for programme compounds (green) overlaid with the training set of known drugs blue, with the candidates are chosen for progression in red.
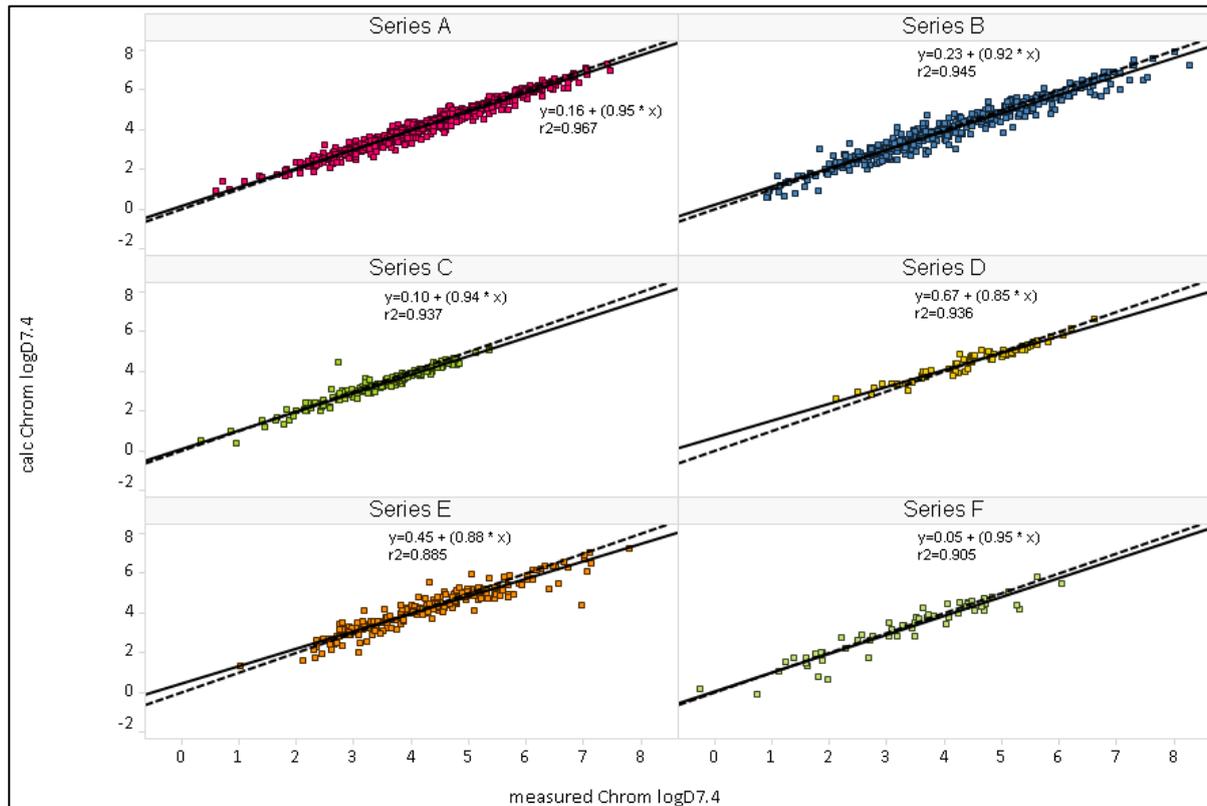
Figure 8. Trellised plot of calculated vs measured Chrom log $D_{7.4}$ for compounds in 6 distinct chemical series with lines of best fit and unity; the r^2 values illustrate the quality of the predictions

Conclusions

The extensive use of high throughput biomimetic measurements impacts on the drug discovery process in many ways. Chromatographic lipophilicity measurements are at the core of Medicinal Chemistry programmes and can be used to predict outcomes, design better compounds and as quality indicators. The complementary measurements from other stationary phases such as HSA and IAM are now routinely used; increasing awareness with demonstrations of their utility and predictive impact. This should give an enhanced influence to programme progression in the future.

References

[1]   R.J. Young, Physical Properties in Drug Design, in: N.A. Meanwell (Ed.) Tactics in Contemporary Drug Design, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 1-68.

[2]   N.A. Meanwell. Improving drug candidates by design: a focus on physicochemical properties as a means of improving compound disposition and safety. *Chemical Research in Toxicology* **24** (2011) 1420-1456.

[3]   M.P. Gleeson, P.D. Leeson, H.van der Waterbeemd, Chapter 1. Physicochemical Properties and Compound Quality, The Handbook of Medicinal Chemistry, The Royal Society of Chemistry2015, pp. 1-31.

[4]   P.D. Leeson, B. Springthorpe. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov* **6** (2007) 881-890.

[5]   P.D. Leeson, R.J. Young. Molecular Property Design: Does Everyone Get It? *ACS Medicinal Chemistry Letters* **6** (2015) 722-725.

[6]   M.P. Gleeson, A. Hersey, D. Montanari, J. Overington. Probing the links between in vitro potency, ADMET and physicochemical parameters. *Nat Rev Drug Discov* **10** (2011) 197-208.

[7]   R.J. Young, P.D. Leeson. Mapping the Efficiency and Physicochemical Trajectories of Successful Optimizations. *Journal of Medicinal Chemistry* **in press** (2018).

[8]   A.P. Hill, R.J. Young. Getting physical in drug discovery: a contemporary perspective on solubility and hydrophobicity. *Drug Discov Today* **15** (2010) 648-655.

[9]   K. Valko, C. Bevan, D. Reynolds. Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log P/log D. *Analytical Chemistry* **69** (1997) 2022-2029.

[10]   R.J. Young, D.V. Green, C.N. Luscombe, A.P. Hill. Getting physical in drug discovery II: the impact of chromatographic hydrophobicity measurements and aromaticity. *Drug Discov Today* **16** (2011) 822-830.

[11]   G. Caron, M. Vallaro, G. Ermondi, G.H. Goetz, Y.A. Abramov, L. Philippe, M. Shalaeva. A Fast Chromatographic Method for Estimating Lipophilicity and Ionization in Nonpolar Membrane-Like Environment. *Molecular Pharmaceutics* **13** (2016) 1100-1110.

[12]   K.L. Valkó. Lipophilicity and biomimetic properties measured by HPLC to support drug discovery. *Journal of Pharmaceutical and Biomedical Analysis* **130** (2016) 35-54.

[13]   M.K. Bayliss, J. Butler, P.L. Feldman, D.V. Green, P.D. Leeson, M.R. Palovich, A.J. Taylor. Quality guidelines for oral drug candidates: dose, solubility and lipophilicity. *Drug Discov Today* **21** (2016) 1719-1727.

[14]   C. Horváth, W. Melander, I. Molnár. Solvophobic interactions in liquid chromatography with nonpolar stationary phases. *Journal of Chromatography A* **125** (1976) 129-156.

[15] F. Lombardo, M.Y. Shalaeva, K.A. Tupper, F. Gao. ElogDoct: A Tool for Lipophilicity Determination in Drug Discovery. 2. Basic and Neutral Compounds. *Journal of Medicinal Chemistry* **44** (2001) 2490-2497.

[16] A. Vailaya, C. Horváth. Solvophobic Theory and Normalized Free Energies of Nonpolar Substances in Reversed Phase Chromatography. *The Journal of Physical Chemistry B* **101** (1997) 5875-5888.

[17] K. Valkó, L.R. Snyder, J.L. Glajch. Retention in reversed-phase liquid chromatography as a function of mobile-phase composition. *Journal of Chromatography A* **656** (1993) 501-520.

[18] M.J. Waring. Lipophilicity in drug discovery. *Expert Opin Drug Discov* **5** (2010) 235-248.

[19] C. Hansch, J.P. Bjorkroth, A. Leo. Hydrophobicity and central nervous system agents: on the principle of minimal hydrophobicity in drug design. *Journal of Pharmaceutical Sciences* **76** (1987) 663-687.

[20] A.L. Hopkins, G.M. Keseru, P.D. Leeson, D.C. Rees, C.H. Reynolds. The role of ligand efficiency metrics in drug discovery. *Nat Rev Drug Discov* **13** (2014) 105-121.

[21] T.J. Ritchie, S.J. Macdonald. Physicochemical descriptors of aromatic character and their use in drug discovery. *Journal of Medicinal Chemistry* **57** (2014) 7206-7215.

[22] T.J. Ritchie, S.J. Macdonald. The impact of aromatic ring count on compound developability--are too many aromatic rings a liability in drug design? *Drug Discov Today* **14** (2009) 1011-1020.

[23] T.J. Ritchie, S.J. Macdonald, R.J. Young, S.D. Pickett. The impact of aromatic ring count on compound developability: further insights by examining carbo- and hetero-aromatic and -aliphatic ring types. *Drug Discov Today* **16** (2011) 164-171.

[24] K. Valko, S. Nunhuck, C. Bevan, M.H. Abraham, D.P. Reynolds. Fast Gradient HPLC Method to Determine Compounds Binding to Human Serum Albumin. Relationships with Octanol/Water and Immobilized Artificial Membrane Lipophilicity. *Journal of Pharmaceutical Sciences* **92** (2003) 2236-2248.

[25] C. Pidgeon, S. Ong, H. Liu, X. Qiu, M. Pidgeon, A.H. Dantzig, J. Munroe, W.J. Hornback, J.S. Kasher. IAM chromatography: an in vitro screen for predicting drug membrane permeability. *Journal of Medicinal Chemistry* **38** (1995) 590-594.

[26] K. Valko, C.M. Du, C.D. Bevan, D.P. Reynolds, M.H. Abraham. Rapid-gradient HPLC method for measuring drug interactions with immobilized artificial membrane: Comparison with other lipophilicity measures. *Journal of Pharmaceutical Sciences* **89** (2000) 1085-1096.

[27] F. Tsopelas, T. Vallianatou, A. Tsantili-Kakoulidou. Advances in immobilized artificial membrane (IAM) chromatography for novel drug discovery. *Expert Opinion on Drug Discovery* **11** (2016) 473-488.

[28] J.-M. Alakoskela, P. Vitovič, P.K.J. Kinnunen. Screening for the Drug–Phospholipid Interaction: Correlation to Phospholipidosis. *ChemMedChem* **4** (2009) 1224-1251.

[29] N. Anderson, J. Borlak. Drug-induced phospholipidosis. *FEBS Letters* **580** (2006) 5533-5540.

[30] S. Braggio, D. Montanari, T. Rossi, E. Ratti. Drug efficiency: a new concept to guide lead optimization programs towards the selection of better clinical candidates. *Expert Opinion on Drug Discovery* **5** (2010) 609-618.

[31] K.L. Valkó, S.B. Nunhuck, A.P. Hill. Estimating Unbound Volume of Distribution and Tissue Binding by In Vitro HPLC-Based Human Serum Albumin and Immobilised Artificial Membrane-Binding Measurements. *Journal of Pharmaceutical Sciences* **100** (2011) 849-862.

[32] K. Valko, E. Chiarparin, S. Nunhuck, D. Montanari. In vitro measurement of drug efficiency index to aid early lead optimization. *Journal of Pharmaceutical Sciences* **101** (2012) 4155-4169.

[33] K.L. Valko, S.P. Teague, C. Pidgeon. In vitro membrane binding and protein binding (IAM MB/PB technology) to estimate in vivo distribution: applications in early drug discovery. *ADMET and DMPK* **5** (2017) 14-38.

[34]   S. Teague, K. Valko. How to identify and eliminate compounds with a risk of high clinical dose during the early phase of lead optimisation in drug discovery. *European Journal of Pharmaceutical Sciences* **110** (2017) 37-50.