Bioorganic & Medicinal Chemistry 17 (2009) 5906–5919

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

ADMET rules of thumb II: A comparison of the effects of common substituents on a range of ADMET parameters

Paul Gleeson [†,*], Gianpaolo Bravi [†], Sandeep Modi [‡], Daniel Lowe [§]

Computational and Structural Chemistry, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY, UK

ARTICLE INFO

Article history:
Received 1 June 2009
Revised 1 July 2009
Accepted 2 July 2009
Available online 7 July 2009

Keywords:
ADMET
SAR
Matched molecular pairs
Solubility
P450 inhibition
hERG inhibition

ABSTRACT

A systematic analysis of data generated in key in vitro assays within GSK has been undertaken to identify what impact a range of common substituents have on a range of ADMET parameters. These include; P450 1A2, 2C9, 2C19, 2D6 and 3A4 inhibition, hERG inhibition, phosphate buffer solubility and artificial membrane permeability. We do this by identifying all matched molecular pairs, differing by the replacement of a hydrogen atom with a list of predefined substituents.

For each substituent we calculate the mean difference in the ADMET parameter for all the matched molecular pairs identified, making a statistical assessment of the difference, as well as assessing the diversity for each example to ensure that the results can be generalized. We also relate the change in activity observed for each substituent to differences in their molecular properties in an effort to identify any structural alerts.

1. Introduction

Approximately 90% of all candidate molecules fail in development due to a variety of causes including; a lack of efficacy, safety and absorption, distribution, metabolism, excretion and toxicity (ADMET) liabilities. This has prompted major effort in early drug discovery to avoid such risks[1] with the development of sophisticated screening cascades.[2] These typically consist of low cost, higher throughput in silico methods at the lead generation (LG) stage, medium throughput in vitro methods during early lead optimization (LO) and lower throughput in vivo methods during late LO.

The introduction of the screening cascade has not reduced the overall attrition rate. However the situation has changed from one where ADME failures dominated in the past,[3] to the current situation where safety and toxicity are the principle causes of concern.[1] It is now generally believed that this situation has arisen due to the fact that the average lipophilicity and molecular weight of molecules now being synthesized are slowly rising.[4–10] This is postulated by some to be due to the rise of combinatorial chemistry and an overemphasis on absolute in vitro potency.[11–14] Even with the judicious selection of low MWT and low lipophilicity, a ligand efficient lead[11,12] is unlikely to have the required potency, selectively and ADMET profile. Thus chemical modifications to the lead will be needed to explore whether it is possible to meet the required candidate profile. In an ideal situation this exercise would be undertaken through the design of a small number of additional molecules by the selection of substituents, or modifications to the core template (scaffold hopping), which might be expected to lead to improved parameters. Knowledge-driven modifications are preferable; however, in many cases suitable data will not exist. This results in most programme teams exploring the 3D space occupied by the lead more extensively with a diverse range of substituents, however this is a more extensive and expensive undertaking.

Knowledge-driven SAR expansion typically relies on the analysis of databases containing relevant biological data and when combined with techniques such as R-group analysis, can be used to directly influence the selection of monomers used in combinatorial libraries to maximize the chances of success. Fragment-based methods are particularly advantageous in SAR expansion exercises as the building blocks used in chemical synthesis can be scored independently and used to determine the next molecules in the series to be synthesized. Commonly used fragment-based Cheminformatics methods include R-group analysis, Free-Wilson analysis,[15] the RECAP[16] technique and fragment-based QSAR models[17] including the popular Daylight *clog P*.[18] An additional technique that is increasingly being used is the matched molecular pair method, where the average effect of a substituent is estimated by analyzing

* Corresponding author at present address: Department of Chemistry, Faculty of Science, Kasetsart University, 50 Phaholyothin Rd, Chatuchak, Bangkok 10900, Thailand. Tel.: +66 (0) 865242120; fax: +66 (0) 25793955.
E-mail address: paul.gleeson@ku.ac.th (P. Gleeson).
† Authors contributed equally.
‡ Present address: Safety & Environmental Assurance Centre, Unilever, Colworth Science Park, Sharnbrook, Bedford, MK44 1LQ, UK.
§ Present address: Unilever Centre for Molecular Science Informatics, University of Cambridge, Department of Chemistry, Lensfield Road Cambridge, CB2 1EW, UK.

0968-0896/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2009.07.002

all the molecular pairs obtained where the only change in structure involves a single, localized substituent change[19–22] (Fig. 1). These methods have been used to assess the mean effect of different substituents on a range of biological parameters that include protein binding,[20,21] solubility,[20,21] bioavailability[21] and primary target activity.[19,22] Unlike the more traditional SAR investigations of congeneric chemical series (local SAR) in the past, these methods are now routinely used to investigate global SAR across a multitude of chemical series, so as to utilize the large amounts of ADMET data available in order to design and synthesize compounds with improved biological and physico-chemical profiles.

Herein we report a wider ranging matched molecular pair study involving the systematic analysis of ~500,000 ADMET datapoints generated in eight in vitro assays commonly run within GlaxoSmithKline (GSK). These assays include cytochrome P450 1A2, 2C9, 2C19, 2D6 and 3A4 inhibition, hERG inhibition, phosphate buffer solubility and artificial membrane permeability.[14] A list of ~90 frequently used substituents has been compiled by our chemists, including halogens, alkyls, amines, alcohols, amides, sulfonamides, esters, acids and a variety of mono- and fused-heterocycles (Supplementary data Table S1). In our matched molecular pairs analysis we search for all pairs of molecules where a hydrogen atom has been replaced by any of these predefined substituents. While this approach will realize more molecular pairs than other comparable studies, giving rise to more reliable statistics, it does have limitations. As noted by one of the referees, the chemical context in which the transformation occurs is likely to play an important role, in that the introduction of an amino group is anticipated to have a drastically different effect depending whether it is added onto an aromatic ring or to an aliphatic chain. Thus, we have specified the chemical context for a short list of frequently occurring substituents where the effect will be noticeable, namely halogens, amines and alcohols. For these we differentiate between the effects of aliphatic and aromatic substitutions.

2. Results

For each of the eight in vitro ADMET datasets studied in this matched molecular pairs analysis, we only report the results for a particular substituent where the total number of pairs obtained is $\geqslant$20 and where these come from $\geqslant$5 daylight fingerprint clusters at a Tanimoto cut-off of 0.7. We subsequently calculate the average difference in the ADMET parameters, *c*log *P*, *c*log *D* and MWT for all the matched pairs obtained for a given substituent. We then make a statistical assessment of the mean differences in the ADMET parameter using Student's *t*-test. The value is considered statistically different from zero if the calculated *P* value is $\leqslant$0.05 (i.e., above the 95% confidence level). In addition, we consider the difference in the means to be of a meaningful size of >0.2 or <0.2 log units.

It is important to consider the experimental variability in the in vitro assays and the numbers of pairs when assessing the activity differences for a given molecular replacement, as a ΔpIC_{50} of +0.2 log units could in fact be highly significant even if the assay standard error is 0.3 log units. in vitro assays typically have two-fold standard errors (0.3 log units) and this tells us that a given measurement on a compound repeated many times will lie between +0.3 and −0.3, 68% of the time, with the mean value centred on zero. Thus for 1000 molecular pairs with a mean difference in activity of +0.3 log units, the differences in the mean will be highly significant, albeit small in magnitude.

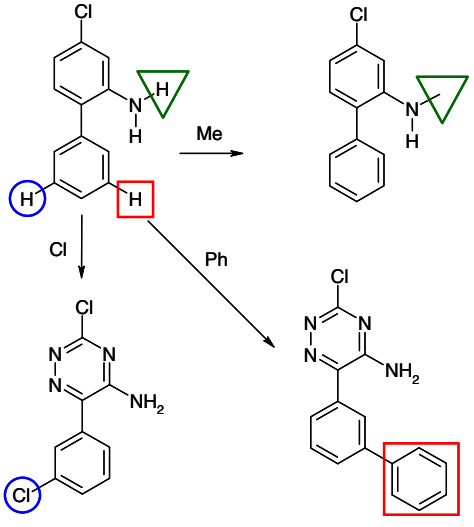


Figure 1. An illustration of the types of molecular changes that are used to estimate the contribution of a range of substituents on a number of key ADMET parameters. All changes involve the replacement of a hydrogen with a predetermined list of commonly used substituents. Where necessary, a distinction between aromatic and aliphatic hydrogen replacements is made, including for amines, hydroxyl, sulfonamides and halides.

2.1. Substitution of hydrogen for a range of common substituents

In Tables 1–8 we report the complete list of results obtained in this investigation. The information reported includes the average change in activity (pIC_{50}, log(Solubility/μM) and log(permeability/nM/s)), molecular weight, *c*log *P* and *c*log *D* (Δ Fragment–Hydrogen). Also reported are the standard deviation in activity (SD), the total numbers of pairs identified for each substituent (N), the *t*-test *P* value and the % of pair changes that lead to an activity change of ±0.2 log units. These tables summarize the effects that are likely to occur when a given substituent is introduced. The interpretation can be straightforward, with many substituents displaying average changes in activities which track closely with their molecular properties. We highlight this effect by considering three different substituents in two of the different in vitro assays. In Figure 2 the results obtained for substituents, hydroxy, ethyl and benzyl are graphically illustrated for P450 3A4 inhibition and phosphate buffer solubility assays. Replacing an aliphatic (Al) hydrogen atom for hydroxyl leads to a average drop in *c*log *D* of −0.9, compared to an increase of +1.2 log units for ethyl and +2.3 for benzyl. Consequently, it is unsurprising that the percentage of pairs where we see a drop in pIC_{50} >0.2 log units are 55%, 15% and 10%, for hydroxyl (Al), ethyl and benzyl, respectively. This effect is also seen for the solubility data, where the same replacements lead to an increase in solubility of +0.2 log units for 48%, 7% and 3% of the pairs for the three substituents. Note that in the case of benzyl substitution, there are three examples where solubility increases were observed. In one case this could be rationalized by the increase in pK_a of the adjacent basic centre while in the other two cases steric crowding may play a role in disrupting the crystal packing.

However there are also exceptions, in that some substituents display changes which are not mirrored by their molecular properties. This is particularly true when a binding event takes place and the shape and size of the binding site pose significant constraints as can be exemplified for the substituent phenyl in the P450 1A2 inhibition assay (Tables 1 and 2). The average change in pIC_{50} calculated for the 147 pairs of molecules identified is as small as +0.11 (P >0.05), in spite of the considerable increase in Δ*c*log *D* of +1.8 log units. This can be contrasted with P450 2C9 and 3A4 inhibition (Tables 5 and 6), where the average ΔpIC_{50} associated with phenyl are +0.38 and +0.46, respectively. In a similar way, the introduction

Table 1
The effect of replacing hydrogen with an alternate substituent on P450 1A2 inhibition

ID	pIC_{50} 1A2 Δ(SD)	%ΔpIC_{50} ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (a)						
Piperidine	−0.68 (0.89)	8; 29; 63	0.34	−1.02	82.98	24 (0.00025)
Dimethylamine (Al)	−0.37 (0.89)	17; 33; 50	−0.42	−1.61	43	24 (0.025)
Morpholine	−0.36 (0.77)	23; 25; 52	−0.65	−0.67	84.89	44 (0.0025)
Acetamide	−0.29 (0.72)	17; 45; 38	−0.8	−0.73	56.72	42 (0.025)
Carboxylic	−0.24 (0.73)	25; 29; 46	−1.23	−2.19	44.01	24 (0.05)
Methyl-alcohol	−0.22 (0.68)	27; 31; 42	−0.83	−0.66	29.84	120 (0.0125)
Amide	−0.20 (0.65)	21; 33; 46	−1.05	−0.67	42.99	56 (0.05)
Cyano	−0.20 (0.65)	25; 33; 42	−0.47	−0.26	24.89	118 (0.0125)
Methylsulfonyl	−0.20 (0.80)	25; 32; 43	−0.63	−0.18	78.06	61 (0.05)
iso-Propyl	−0.19 (0.72)	30; 29; 41	1.29	1.48	42.06	79 (0.025)
Methyl-ester	−0.17 (0.55)	18; 44; 38	−0.04	0.27	57.99	40 (0.05)
Acetyl	−0.17 (0.68)	20; 43; 37	0.04	0.35	42.04	75 (0.05)
Methylrevsulfonamide	−0.15 (0.88)	13; 45; 42	−1.3	−1.32	92.86	24 (0.2)
Hydroxy (Al)	−0.12 (0.55)	22; 37; 41	−1.1	−0.88	15.91	181 (0.025)
Ethoxy	−0.12 (0.46)	17; 53; 30	0.34	0.38	44.06	23 (0.125)*
Benzyl	−0.10 (0.89)	43; 19; 38	2.06	2.17	90.2	60 (0.125)*
Trifluoromethoxy	−0.06 (0.58)	40; 32; 28	1.07	1	83.52	25 (0.2)
Methoxy	−0.05 (0.54)	26; 41; 33	−0.09	−0.03	29.9	467 (0.125)
Dimethylether	−0.04 (0.44)	24; 48; 28	0.09	0.1	43.81	89 (0.2)
Methyl	−0.04 (0.51)	28; 41; 31	0.4	0.48	13.97	2459 (0.005)
Ethyl	−0.04 (0.58)	32; 36; 32	1.01	1.13	27.96	260 (0.125)
Fluoro (Ar)	−0.03 (0.45)	26; 42; 32	0.13	0.12	17.9	1244 (0.05)
Propyl	−0.02 (0.64)	37; 29; 34	1.48	1.55	41.99	62 (0.5)
Trifluoromethyl	0.03 (0.55)	32; 38; 30	0.89	1.11	67.88	179 (0.2)
n-Butyl	0.04 (0.72)	35; 48; 17	2.01	2.1	55.91	29 (0.5)
Dimethylamine (Ar)	0.08 (0.78)	38; 21; 41	0.29	0.17	42.66	29 (0.2)
Furan	0.09 (0.56)	33; 42; 25	1.02	1.15	66.06	24 (0.2)
Chloro (Ar)	0.10 (0.51)	40; 37; 23	0.66	0.71	34.3	800 (0.0025)
Pyridyl	0.11 (0.84)	43; 26; 31	0.5	0.51	77.09	58 (0.125)*
Phenyl	0.11 (0.73)	46; 27; 27	1.77	1.84	76.11	147 (0.05)
Bromo (Ar)	0.12 (0.47)	41; 37; 22	0.87	0.91	78.83	144 (0.05)
t-Butyl	0.13 (0.41)	39; 35; 26	1.71	2.03	56.12	23 (0.125)*
Thiophene	0.25 (0.59)	55; 22; 23	1.57	1.93	82.03	22 (0.125)*

Reported are the mean changes in pIC_{50}, molecular weight and clog *P* (Δ Fragment–Hydrogen). Also reported is the standard deviation in pIC_{50} (SD) and the total numbers of pairs (*N*) for each mutation, along with the *t*-test *P* value and the % of pair changes that lead to an increase or decrease in pIC_{50} of +/−0.2. The substituents are sorted according to their activity. Where the change in pIC_{50} is >+/−0.1 but one can only be <95% confident in the result it is denoted with a *.

Table 2
The effect of replacing hydrogen with an alternate substituent on 2C9 inhibition

ID	pIC_{50} 2C9 Δ(SD)	%ΔpIC_{50} ⩾−0.2;**<>**;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (b)						
Carboxylic	−0.76 (0.80)	15; 8; 77	−0.68	−2.64	43.93	53 (0.00025)
Piperidine	−0.38 (1.01)	27; 23; 50	0.51	−0.6	82.88	30 (0.025)
Amine (Al)	−0.37 (0.56)	14; 26; 60	−1.41	−2.55	14.53	37 (0.005)
Dimethylamine (Al)	−0.32 (0.60)	9; 29; 62	−0.37	−1.18	42.9	34 (0.025)
Amide	−0.15 (0.59)	22; 25; 53	−1.34	−1.13	42.71	81 (0.05)
Hydroxy (Al)	−0.15 (0.43)	18; 33; 49	−1.14	−0.99	15.97	318 (0.0025)
Methyl-alcohol	−0.07 (0.44)	24; 33; 43	−0.71	−0.58	29.87	241 (0.05)
Acetamide	−0.04 (0.62)	29; 33; 38	−0.8	−0.75	56.9	86 (0.2)
Methylrevsulfonamide	−0.02 (0.45)	30; 29; 41	−1.29	−1.36	92.93	46 (0.5)
Morpholine	−0.01 (0.55)	23; 42; 35	−0.53	−0.97	84.9	65 (0.5)
Methoxy	0.02 (0.48)	34; 38; 28	−0.1	−0.02	29.85	966 (0.2)
Aro-fluoro	0.04 (0.37)	30; 47; 23	0.13	0.1	17.87	2683 (0.0025)
Cyano	0.05 (0.51)	35; 32; 33	−0.48	−0.29	24.79	272 (0.125)
Methyl	0.09 (0.43)	37; 42; 21	0.39	0.47	13.95	4685 (0.00025)
Acetyl	0.10 (0.60)	32; 37; 31	−0.14	0.27	42.04	129 (0.125)*
t-Butyl	0.10 (0.67)	42; 23; 35	1.78	1.7	55.93	65 (0.125)
Fluoro-methyl	0.12 (0.31)	36; 46; 18	0.25	0.41	32.02	22 (0.2)
Nitro	0.14 (0.44)	53; 27; 20	0.03	0.31	44.93	30 (0.125)*
Ethoxy	0.15 (0.50)	50; 30; 20	0.47	0.5	44.06	54 (0.05)
Methylsulfonyl	0.17 (0.56)	36; 41; 23	−0.81	−0.49	77.98	107 (0.0125)
Dimethylether	0.19 (0.49)	46; 34; 20	0.06	0.04	43.8	158 (0.005)
Difluoro-methyl	0.19 (0.51)	48; 33; 19	0.51	0.49	50.01	21 (0.125)*
Chloro (Ar)	0.19 (0.45)	49; 34; 17	0.66	0.71	34.28	1550 (0.00025)
Furan	0.19 (0.65)	51; 22; 27	1.06	1.21	66.06	37 (0.05)
iso-Propyl-oxy	0.22 (0.45)	57; 19; 24	0.77	0.9	57.51	21 (0.125)*
Ethyl	0.22 (0.48)	53; 29; 18	1.06	1.2	27.96	539 (0.00025)
Dimethylamine (Ar)	0.23 (0.44)	59; 28; 13	0.26	0.07	42.51	46 (0.025)
Ethene	0.23 (0.32)	52; 40; 8	0.74	0.88	26.04	25 (0.125)*
n-Butyl	0.23 (0.62)	58; 23; 19	1.94	2.13	56.1	43 (0.025)
Bromo (Ar)	0.24 (0.50)	57; 27; 16	0.87	0.97	78.83	299 (0.00025)
Thiomethyl	0.25 (0.51)	50; 33; 17	0.5	0.61	46.09	24 (0.05)
Thiophene	0.25 (0.78)	58; 13; 29	1.61	1.76	82.03	45 (0.025)
Trifluoromethyl	0.26 (0.54)	52; 31; 17	0.9	1.04	67.9	514 (0.00025)
iso-Propyl	0.26 (0.60)	55; 24; 21	1.36	1.47	41.99	167 (0.00025)
Cyclo-hexyl	0.27 (0.86)	50; 15; 35	2.43	2.62	81.86	20 (0.125)*
i-Butyl	0.28 (0.75)	59; 23; 18	1.85	1.92	56.12	44 (0.025)
Trifluoromethoxy	0.29 (0.64)	52; 29; 19	1.03	1.05	83.87	62 (0.0125)
Methyl-ester	0.30 (0.54)	55; 31; 14	0	0.34	57.7	65 (0.005)
Propyl	0.37 (0.46)	65; 22; 13	1.52	1.61	42.03	109 (0.00025)
Phenyl	0.38 (0.69)	61; 19; 20	1.84	1.91	76.18	290 (0.00025)
Pyridyl	0.46 (0.67)	68; 17; 15	0.47	0.66	77.09	101 (0.00025)
Thiazole	0.49 (0.84)	58; 13; 29	0.44	0.65	83.11	24 (0.0025)
Benzyl	0.68 (0.63)	80; 11; 9	2.07	2.23	90.08	112 (0.00025)
Pyrimidine	0.79 (0.88)	67; 25; 8	0.16	0.66	78.08	24 (0.00025)

Table 3
The effect of replacing hydrogen with an alternate substituent on P450-2C19 inhibition

ID	pIC_{50} 2C19 Δ(SD)	%ΔpIC_{50} ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (a)						
Amide	−0.41 (0.70)	16; 23; 61	−1.14	−0.84	42.95	49 (0.00025)
Carboxylic	−0.27 (0.64)	22; 21; 57	−0.93	−2.47	43.92	23 (0.05)
Hydroxy (Al)	−0.18 (0.44)	14; 34; 52	−1.13	−0.98	15.98	230 (0.00025)
Methyl-alcohol	−0.16 (0.46)	13; 39; 48	−0.73	−0.61	29.98	157 (0.0025)
Amine (Al)	−0.09 (0.61)	28; 28; 44	−1.38	−2.62	14.89	32 (0.125)
Acetamide	−0.08 (0.52)	25; 36; 39	−0.77	−0.8	56.77	49 (0.2)
Methyl-ester	−0.04 (0.57)	31; 36; 33	−0.03	0.47	57.56	42 (0.2)
Morpholine	−0.02 (0.72)	32; 25; 43	−0.5	−0.75	84.8	44 (0.5)
Cyano	−0.02 (0.48)	28; 37; 35	−0.5	−0.29	24.81	200 (0.2)
Acetyl	−0.02 (0.54)	23; 48; 29	−0.11	0.33	42.04	92 (0.5)
Ethoxy	0.02 (0.35)	31; 43; 26	0.59	0.59	44.06	35 (0.5)
Methylsulfonyl	0.04 (0.59)	32; 36; 32	−0.83	−0.49	78.01	75 (0.2)
Fluoro (Ar)	0.04 (0.37)	30; 48; 22	0.13	0.1	17.87	1860 (0.025)
Methoxy	0.05 (0.41)	31; 46; 23	−0.13	−0.07	29.83	663 (0.05)
Fluoro-methyl	0.05 (0.25)	24; 62; 14	0.26	0.39	32.02	21 (0.2)
Methyl	0.05 (0.43)	33; 46; 21	0.39	0.48	13.97	3386 (0.00025)
Methylrevsulfonamide	0.09 (0.53)	44; 28; 28	−1.24	−1.27	93.35	25 (0.2)
Furan	0.11 (0.50)	44; 28; 28	1.06	1.21	66	32 (0.125)*
Piperidine	0.12 (0.80)	50; 14; 36	0.68	−0.56	82.93	28 (0.125)*
Dimethylether	0.12 (0.45)	39; 42; 19	0.03	−0.01	43.78	128 (0.025)
Chloro (Ar)	0.13 (0.48)	43; 36; 21	0.66	0.71	34.3	1085 (0.00025)
Dimethylamine (Al)	0.14 (0.75)	33; 43; 24	−0.38	−1.32	42.98	21 (0.125)*
iso-Propyl	0.16 (0.65)	46; 29; 25	1.31	1.5	42.11	118 (0.0125)
Thiophene	0.17 (0.67)	44; 26; 30	1.68	2.04	82.12	23 (0.125)
Trifluoromethoxy	0.18 (0.49)	54; 21; 25	1.04	1.01	83.89	56 (0.025)
Trifluoromethyl	0.18 (0.49)	51; 29; 20	0.85	1.05	67.84	301 (0.00025)
Ethyl	0.18 (0.45)	48; 34; 18	1.01	1.17	27.98	393 (0.00025)
Dimethylamine (Ar)	0.19 (0.56)	50; 35; 15	0.35	0.15	42.78	34 (0.05)
Propyl	0.26 (0.67)	54; 21; 25	1.51	1.56	42.04	87 (0.0025)
Bromo (Ar)	0.27 (0.50)	54; 32; 14	0.88	0.92	78.82	195 (0.00025)
t-Butyl	0.30 (0.60)	54; 23; 23	1.77	1.72	56.12	35 (0.005)
i-Butyl	0.37 (0.73)	58; 19; 23	1.81	1.76	55.92	31 (0.005)
Phenyl	0.38 (0.56)	64; 23; 13	1.82	1.86	76.14	188 (0.00025)
n-Butyl	0.44 (0.67)	64; 11; 25	1.91	2.06	56.04	36 (0.00025)
Benzyl	0.47 (0.58)	72; 15; 13	2.11	2.22	90.25	68 (0.00025)
Pyridyl	0.48 (0.64)	63; 21; 16	0.51	0.78	77.09	80 (0.00025)
Pyrimidine	0.59 (0.95)	60; 10; 30	−0.08	0.65	78.08	20 (0.0125)

See Table 1 legend for specific details.

Table 4
The effect of replacing hydrogen with an alternate substituent on 2D6 inhibition

ID	pIC_{50} 2D6 Δ(SD)	% ΔpIC_{50} ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (b)						
Carboxylic	−0.67 (0.70)	8; 18; 74	−1.81	−1.61	44.13	50 (0.00025)
Dimethylamine (Ar)	−0.26 (0.66)	24; 33; 43	0.35	0.17	42.92	37 (0.025)
Amine (Al)	−0.23 (0.75)	24; 30; 46	−1.26	−2.42	14.87	41 (0.05)
Methylsulfonyl	−0.21 (0.77)	26; 28; 46	−0.7	−0.05	78.15	70 (0.025)
Methyl-ester	−0.21 (0.60)	23; 26; 51	0.04	0.51	57.79	65 (0.025)
Amide	−0.20 (0.62)	27; 25; 48	−1.31	−1.04	43.07	102 (0.0125)
Hydroxy (Al)	−0.16 (0.52)	21; 30; 49	−1.08	−0.88	15.86	323 (0.0025)
Dimethylamine (Al)	−0.15 (0.64)	32; 33; 35	−0.28	−1.31	43.02	34 (0.125)*
Cyano	−0.11 (0.55)	26; 35; 39	−0.55	−0.29	24.89	297 (0.025)
Ethoxy	−0.11 (0.61)	29; 29; 42	0.49	0.59	44.02	48 (0.125)*
Methoxy	−0.08 (0.50)	25; 39; 36	−0.13	−0.04	29.86	813 (0.005)
Methyl-alcohol	−0.07 (0.48)	25; 37; 38	−0.71	−0.49	29.92	191 (0.125)
Acetyl	−0.07 (0.55)	33; 30; 37	−0.21	0.46	41.95	110 (0.125)
Piperidine	−0.03 (0.72)	47; 24; 29	0.48	−0.74	82.97	34 (0.5)
Morpholine	−0.02 (0.59)	36; 33; 31	−0.49	−0.65	84.99	45 (0.5)
Methyl	0.02 (0.49)	32; 41; 27	0.36	0.47	13.95	3883 (0.125)
Aro-fluoro	0.03 (0.41)	30; 47; 23	0.13	0.12	17.87	2087 (0.05)
Dimethylether	0.03 (0.37)	30; 43; 27	0.07	0.16	43.85	145 (0.2)
Methylrevsulfonamide	0.04 (0.50)	41; 26; 33	−1.25	−1.27	93.07	46 (0.2)
Trifluoromethyl	0.06 (0.56)	42; 33; 25	0.88	1.07	67.79	327 (0.05)
Acetamide	0.07 (0.58)	40; 30; 30	−1.01	−0.98	56.87	73 (0.2)
Chloro (Ar)	0.09 (0.48)	40; 37; 23	0.67	0.73	34.26	1238 (0.00025)
t-Butyl	0.11 (0.65)	43; 31; 26	1.85	1.8	55.65	47 (0.125)
Nitro	0.14 (0.62)	55; 22; 23	0	0.35	45	22 (0.2)
Trifluoromethoxy	0.17 (0.67)	45; 34; 21	1.03	1.07	83.76	58 (0.05)
Bromo (Ar)	0.18 (0.53)	48; 31; 21	0.84	0.94	78.78	232 (0.0025)
Phenyl	0.18 (0.80)	49; 23; 28	1.79	1.87	76.02	256 (0.0025)
Ethyl	0.24 (0.57)	51; 29; 20	1.24	1.23	27.98	512 (0.00025)
Thiophene	0.26 (0.70)	48; 35; 17	1.58	1.95	82.12	29 (0.125)
iso-Propyl	0.28 (0.62)	56; 25; 19	1.33	1.59	42.06	158 (0.00025)
n-Butyl	0.28 (0.60)	48; 26; 26	1.83	2.02	55.63	27 (0.125)*
Furan	0.31 (0.66)	55; 24; 21	0.89	1.25	66.24	33 (0.025)
Pyridyl	0.36 (0.71)	68; 12; 20	0.6	0.91	77.17	74 (0.0025)
Propyl	0.36 (0.60)	63; 21; 16	1.49	1.63	41.99	80 (0.0025)
i-Butyl	0.37 (0.59)	58; 21; 21	1.84	1.86	55.96	38 (0.005)
Benzyl	0.40 (0.76)	60; 20; 20	2.05	2.28	89.89	98 (0.00025)

See Table 1 legend for specific details.

Table 5
The effect of replacing hydrogen with an alternate substituent on P450-3A4 inhibition

ID	pIC_{50} 3A4 Δ(SD)	%ΔpIC_{50} ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (a)						
Carboxylic	−0.55 (0.56)	8; 14; 78	−1.71	−1.75	44.09	50 (0.00025)
Hydroxy (Al)	−0.23 (0.55)	20; 25; 55	−1	−0.9	15.93	389 (0.00025)
Methyl-alcohol	−0.05 (0.49)	32; 31; 37	−0.61	−0.46	29.94	225 (0.125)
Amide	−0.01 (0.52)	33; 30; 37	−1.24	−0.88	42.87	112 (0.5)
Piperidine	0.00 (0.90)	42; 19; 39	0.39	−0.99	82.56	31 (0.5)
Amine (Al)	0.07 (0.53)	53; 17; 30	−1.26	−2.25	14.64	40 (0.2)
Fluoro (Ar)	0.07 (0.42)	33; 46; 21	0.12	0.1	17.86	2297 (0.00025)
Cyano	0.08 (0.44)	30; 51; 19	−0.57	−0.29	24.86	275 (0.025)
Dimethylamine (Al)	0.09 (0.74)	50; 25; 25	−0.31	−1.19	43	36 (0.2)
Methoxy	0.11 (0.48)	39; 40; 21	−0.11	−0.06	29.91	785 (0.00025)
Methyl	0.11 (0.50)	41; 37; 22	0.38	0.47	13.96	4518 (0.00025)
Methylsulfonyl	0.12 (0.66)	42; 21; 37	−0.6	−0.07	78	95 (0.125)*
Acetyl	0.16 (0.66)	47; 27; 26	−0.13	0.69	42.06	114 (0.025)
Dimethylether	0.19 (0.46)	42; 39; 19	0.09	0.11	43.88	178 (0.0025)
n-Butyl	0.20 (0.75)	44; 26; 30	1.97	2.09	55.97	46 (0.05)
Chloro (Ar)	0.21 (0.44)	51; 35; 14	0.66	0.7	34.35	1225 (0.00025)
Ethoxy	0.22 (0.66)	53; 30; 17	0.52	0.6	43.93	47 (0.05)
Acetamide	0.23 (0.54)	49; 30; 21	−1.1	−0.98	56.52	78 (0.0125)
Thiomethyl	0.24 (0.39)	43; 47; 10	0.55	0.82	46.19	21 (0.125)
Trifluoromethyl	0.24 (0.51)	56; 25; 19	0.83	1.1	67.83	353 (0.00025)
Methyl-ester	0.26 (0.53)	47; 33; 20	−0.04	0.52	57.8	51 (0.025)
Bromo (Ar)	0.27 (0.48)	59; 28; 13	0.84	0.92	78.76	203 (0.00025)
Methylrevsulfonamide	0.28 (0.56)	57; 26; 17	−1.31	−1.26	92.39	42 (0.0125)
Propyl	0.28 (0.58)	55; 27; 18	1.49	1.6	41.97	101 (0.0025)
iso-Propyl	0.29 (0.62)	55; 27; 18	1.32	1.56	42.1	166 (0.00025)
Ethyl	0.32 (0.56)	60; 25; 15	1.18	1.26	27.99	505 (0.00025)
Dimethylamine (Ar)	0.36 (0.63)	68; 11; 21	0.38	0.09	42.44	38 (0.005)
t-Butyl	0.36 (0.68)	61; 21; 18	1.87	1.76	55.91	49 (0.0025)
Morpholine	0.42 (0.69)	67; 13; 20	−0.55	−0.68	84.68	55 (0.00025)
Trifluoromethoxy	0.44 (0.46)	70; 23; 7	1.01	1.06	83.96	46 (0.00025)
Phenyl	0.46 (0.54)	67; 20; 13	1.77	1.89	76.06	257 (0.00025)
i-Butyl	0.50 (0.62)	61; 25; 14	1.85	1.78	56.12	43 (0.00025)
Cyclohexyl	0.51 (0.61)	75; 8; 17	2.52	2.62	82.16	24 (0.005)
Fluoro-methyl	0.52 (0.72)	55; 27; 18	0.26	0.69	31.93	22 (0.0125)
Furan	0.54 (0.62)	79; 7; 14	0.9	1.33	66.28	28 (0.0025)
Benzyl	0.65 (0.68)	78; 12; 10	2.13	2.4	90.13	106 (0.00025)
Thiophene	0.73 (0.66)	80; 9; 11	1.41	1.86	81.83	35 (0.00025)
Pyridyl	0.86 (0.83)	86; 6; 8	0.5	0.82	77.01	78 (0.00025)

See Table 1 legend for specific details.

Table 6
The effect of replacing hydrogen with an alternate substituent on hERG inhibition

ID	pIC_{50} hERG Δ(SD)	%ΔpIC_{50} ⩾−0.2;**<>**;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (b)						
Amide	−0.52 (0.69)	13; 12; 75	−1.19	−0.99	42.65	69 (0.00025)
Methylsulfonyl	−0.50 (0.68)	13; 16; 71	−1.15	−1.05	78.06	68 (0.00025)
Acetyl	−0.39 (0.64)	22; 13; 65	−0.19	0.24	42.04	37 (0.0025)
Dimethylamine (Al)	−0.35 (0.63)	13; 30; 57	−0.68	−1.49	42.73	23 (0.005)
Acetamide	−0.34 (0.69)	21; 25; 54	−0.96	−0.74	56.92	28 (0.025)
Hydroxy (Al)	−0.33 (0.52)	16; 20; 64	−1.07	−0.74	15.88	196 (0.00025)
Methyl-ester	−0.26 (0.45)	27; 20; 53	−0.04	0.4	57.86	34 (0.025)
Methyl-alcohol	−0.09 (0.49)	28; 32; 40	−0.58	−0.27	29.9	95 (0.125)
iso-Propyl-oxy	−0.01 (0.51)	33; 34; 33	0.29	0.53	57.8	21 (0.5)
Dimethylether	0.00 (0.50)	36; 25; 39	0.03	−0.02	43.75	59 (0.5)
Methoxy	0.05 (0.56)	42; 22; 36	−0.13	−0.04	29.91	644 (0.125)
Methyl	0.08 (0.52)	45; 26; 29	0.3	0.44	13.97	2398 (0.00025)
Fluoro (Ar)	0.09 (0.49)	43; 30; 27	0.14	0.2	17.89	1029 (0.0025)
Cyano	0.13 (0.61)	46; 22; 32	−0.63	−0.32	24.85	191 (0.125)*
Pyridyl	0.13 (0.66)	48; 25; 27	0.41	0.74	77.04	44 (0.125)*
Chloro (Ar)	0.23 (0.53)	57; 22; 21	0.64	0.73	34.27	837 (0.00025)
Trifluoromethyl	0.27 (0.70)	58; 14; 28	0.88	1.07	67.74	247 (0.00025)
Ethoxy	0.30 (0.60)	57; 25; 18	0.1	0.39	43.92	28 (0.05)
iso-Propyl	0.30 (0.50)	61; 19; 20	1.27	1.52	41.95	147 (0.00025)
Nitro	0.33 (0.72)	61; 17; 22	−0.15	−0.04	45	23 (0.05)
Ethene	0.33 (0.58)	55; 31; 14	0.69	0.69	26.04	22 (0.125)*
Ethyl	0.33 (0.60)	61; 19; 20	1.02	1.14	27.99	242 (0.00025)
Bromo (Ar)	0.35 (0.55)	64; 17; 19	0.79	0.83	78.71	167 (0.00025)
Trifluoromethoxy	0.39 (0.63)	60; 25; 15	1.01	1.04	83.85	40 (0.005)
t-Butyl	0.53 (0.44)	74; 20; 6	1.89	1.74	56.06	35 (0.00025)
Phenyl	0.56 (0.71)	72; 13; 15	1.73	1.92	75.89	326 (0.00025)
Propyl	0.68 (0.68)	75; 13; 12	1.5	1.55	42.04	89 (0.00025)
Benzyl	0.70 (0.94)	70; 8; 22	2	2.45	89.92	171 (0.00025)
n-Butyl	1.06 (0.77)	87; 6; 7	1.67	1.74	55.96	62 (0.00025)

See Table 1 legend for specific details. See text regarding the hERG value for di-methyl-amine (Al).

Table 7
The effect of replacing hydrogen with an alternate substituent on Solubility (log (mM/L))

ID	logSOL Δ(SD)	%ΔlogSOL ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (a)						
Trifluoromethoxy	−0.72 (0.51)	4; 12; 84	1.04	0.97	84	25 (0.00025)
Furan	−0.65 (0.58)	4; 25; 71	1.07	0	66.06	24 (0.00025)
Thiophene	−0.6 (0.64)	3; 37; 60	1.47	2.03	81.78	30 (0.00025)
Phenyl	−0.58 (0.59)	3; 29; 68	1.72	1.91	75.98	171 (0.00025)
Trifluoromethyl	−0.54 (0.52)	3; 25; 72	0.86	0.94	67.97	245 (0.00025)
Bromo (Ar)	−0.53 (0.52)	2; 27; 71	0.85	0.81	78.79	166 (0.00025)
Benzyl	−0.49 (0.52)	3; 31; 66	2.09	2.25	89.95	91 (0.00025)
t-Butyl	−0.47 (0.52)	0; 43; 57	1.83	0	56.12	28 (0.00025)
Propyl	−0.4 (0.48)	1; 31; 68	1.46	1.41	42.03	69 (0.00025)
Chloro (Ar)	−0.35 (0.47)	5; 44; 51	0.67	0.59	34.4	892 (0.00025)
n-Butyl	−0.35 (0.51)	8; 36; 56	1.92	1.92	56.04	25 (0.0025)
Ethyl	−0.3 (0.45)	7; 44; 49	0.93	1.02	27.97	420 (0.00025)
i-Butyl	−0.3 (0.43)	17; 29; 54	1.89	1.7	56.12	24 (0.0125)
Ethene	−0.29 (0.44)	7; 41; 52	0.73	0.8	26.04	29 (0.0125)
Pyridyl	−0.25 (0.39)	3; 59; 38	0.32	0.48	76.95	134 (0.00025)
Ethoxy	−0.2 (0.49)	16; 35; 49	0.33	0.44	44.06	37 (0.0125)
Dimethylamine (Ar)	−0.19 (0.48)	11; 51; 38	0.22	0.02	43.08	37 (0.05)
Cyano	−0.14 (0.47)	9; 64; 27	−0.49	−0.37	24.91	170 (0.005)
Methyl-ester	−0.14 (0.39)	11; 57; 32	0.07	0.27	58	57 (0.025)
Methyl	−0.11 (0.4)	11; 59; 30	0.36	0.43	13.96	3356 (0.00025)
Fluoro (Ar)	−0.1 (0.34)	9; 67; 24	0.13	0.11	17.93	1572 (0.00025)
Difluoro-methyl	−0.07 (0.42)	22; 52; 26	0.31	0.11	50.01	23 (0.2)
Acetyl	−0.05 (0.47)	18; 60; 22	−0.12	0.65	42.04	107 (0.125)
Methoxy	−0.03 (0.44)	19; 57; 24	−0.08	−0.1	29.99	642 (0.125)
Dimethylether	0 (0.28)	13; 68; 19	0.14	0.05	43.94	100 (0.5)
Methylsulfonyl	0.01 (0.6)	27; 50; 23	−0.94	−0.41	78.04	117 (0.5)
Morpholine	0.06 (0.5)	25; 57; 18	−0.53	−0.46	85.07	44 (0.2)
Dimethylamine (Al)	0.11 (0.5)	26; 63; 11	−0.45	−0.63	43.03	38 (0.125)
Amide	0.15 (0.5)	34; 54; 12	−1.13	−0.69	43	121 (0.005)
Piperidine	0.24 (0.71)	40; 40; 20	0.61	1.2	83.05	40 (0.0125)
Acetamide	0.25 (0.53)	52; 33; 15	−0.79	−0.61	57.06	60 (0.0025)
Methylrevsulfonamide	0.27 (0.47)	44; 52; 4	−1.14	−1.2	93.07	48 (0.0025)
Methyl-alcohol	0.28 (0.45)	38; 56; 6	−0.64	−0.6	29.94	149 (0.00025)
Hydroxy (Al)	0.31 (0.5)	48; 44; 8	−1.05	−1.01	15.86	172 (0.00025)
Amine (Al)	0.37 (0.62)	54; 38; 8	−1.05	−3.13	14.85	24 (0.0025)
Carboxylic	0.57 (0.65)	56; 40; 4	−0.97	−2.7	44.01	54 (0.00025)

See Table 1 legend for specific details.

of a furan leads to average pIC_{50} increases across P450 1A2, 2C9 and 3A4 inhibition of +0.09, +0.19 and +0.54, respectively. These trends may presumably be related to the cavity sizes of the three P450 isoforms, with the 3A4 cavity being significantly larger than 2C9, and the 1A2 cavity considerably smaller.

The values reported in Tables 1–8 represent an average effect of a given substituent. However, the average effect can mask the subtle differences when considering all the possible variations of a given substituent. This is the case of pyridyl, for example, and its regioisomers: 2-, 3- and 4-pyridyl. The individual pair comparisons obtained for pyridyl from the P450 2C9 dataset are illustrated in Figure 3a, displaying a distinct shift to higher pIC_{50}s when hydrogen is substituted for pyridyl. The mean difference is +0.46, which is a statistically significant difference at the 95% confidence level (P >0.00025), with only 15% of the 15 cases leading to a drop in pIC_{50} below 0.2 log units (Tables 1 and 2). Breaking the data down further reveals considerably different results depending on the position of the nitrogen atom in the pyridyl ring (Fig. 3b). For example, the most accessible lone pair is found in 4-pyridyl and it shows demonstrably greater liability than 2- or 3-substituted pyridyl, presumably on account of the nitrogen lone pair being capable of coordinating more effectively with haem group. 40% of all 4-pyridyls lead to an increase in activity of >1 log unit when substituted for hydrogen, in contrast to 13% for 3-pyridyl and just 6% for 2-pyridyl.

Pyrimidine is another example which displays significant inhibition at P450 2C9, leading to an even larger ΔpIC_{50} than pyridyl (+0.79 vs +0.46). However, the 2C9 ΔpIC_{50} standard deviation (SD) for pyrimidine is high at +0.88 (Tables 1 and 2) and from the individual example reported in Table 9 it is clear that the effect of substituting pyrimidine for hydrogen can be very different depending on where substitution occurs. For three different chemotypes we show two different examples to ensure that the effect is real and not a result of assay error. In cases A and B, substitution of an aliphatic hydrogen for pyrimidine leads to an average ΔpIC_{50} of −0.15 and a drop in *c*log *D* of −0.75. Given that the average ΔpIC_{50} is +0.79 for the 24 pyrimidine pairs obtained for 2C9, with a mean Δclog *D* being +0.7, this suggests that the substituent in A2 and B2 may not chelate to the haem. Thus, the observed change in pIC_{50} is primarily a result of the *c*log *D* change. This can be contrasted with pairs C and D, where we observe a mean ΔpIC_{50} of +0.45 and a drop in log *D* of −0.6. The increase in 2C9 inhibition despite the drop in *c*log *D* could be due to an undesirable chelation of the haem. The final set of pairs (E and F) display a different behaviour again, since the substitution involves the replacement of a secondary amine hydrogen for pyrimidine. This leads to a dramatic reduction in the basic pK_a of the amine and an increase in *c*log *D* of +2.4 log units. This also leads to a dramatic change in the average pIC_{50} (+2.3), with molecules E2 and F2 having low μM affinities. In this case the dramatic change could be down to the lipophilicity change alone, although one cannot exclude the possibility that haem chelation occurs.

Tables 1–8 summarize the average effect of introducing common chemical groups across a number of ADMET assays. They have been derived considering a wide range of chemotypes studied within GSK over many years, paying attention to exclude substituents where the underlying diversity of the set is low so as to minimize issues about generalizability. Nonetheless, for the reasons

Table 8
The effect of replacing hydrogen with an alternate substituent on Permeability (log (nm/s))

ID	logPerm Δ(SD)	% ΔlogPerm ⩾−0.2;<>;⩽0.2	Δclog *P*	Δclog *D*	ΔMWT	*N* (>*P*)
Panel (b)						
Carboxylic	−1.19 (0.9)	4; 9; 87	−0.83	−1.72	43.95	139 (0.00025)
Amine (Al)	−0.91 (0.73)	5; 13; 82	−1.18	−1.66	14.9	65 (0.00025)
Acetamide	−0.72 (0.72)	10; 17; 73	−1.02	−0.95	57	109 (0.00025)
Methyl revsulfonamide	−0.6 (0.7)	15; 16; 69	−1.13	−1.2	93.01	39 (0.00025)
Sulfonamide	−0.54 (0.6)	3; 28; 69	−1.01	−0.66	79.08	68 (0.00025)
Amide	−0.52 (0.63)	8; 27; 65	−1.03	−0.88	42.94	226 (0.00025)
Hydroxy (Al)	−0.38 (0.57)	10; 34; 56	−0.96	−0.76	15.91	596 (0.00025)
Methyl-alcohol	−0.33 (0.53)	13; 34; 53	−0.81	−0.3	29.91	360 (0.00025)
Methylamine	−0.26 (0.57)	11; 43; 46	0.06	−1.34	28.83	28 (0.025)
Dimethylamine (Al)	−0.19 (0.6)	19; 46; 35	−0.18	−1.22	42.75	92 (0.005)
Piperidine	−0.18 (0.67)	18; 48; 34	0.01	−0.74	83	68 (0.125)
Nitro	−0.17 (0.44)	8; 60; 32	−0.09	0.42	44.83	165 (0.0025)
Morpholine	−0.16 (0.43)	17; 40; 43	−0.66	−0.56	85.02	117 (0.05)
Methylsulfonyl	−0.13 (0.65)	24; 39; 37	−0.69	−0.11	78.03	244 (0.005)
Cyano	−0.11 (0.44)	15; 53; 32	−0.48	−0.27	24.89	497 (0.0025)
Acetyl	−0.1 (0.54)	20; 42; 38	−0.24	0.57	41.97	236 (0.025)
Trifluoromethyl	−0.07 (0.43)	20; 46; 34	0.94	1	67.88	755 (0.0025)
Trifluoromethoxy	−0.07 (0.44)	21; 47; 32	1.06	1.11	83.57	75 (0.125)
Ethene	−0.05 (0.23)	12; 63; 25	0.63	0.93	25.89	97 (0.125)
Dimethylamine (Ar)	−0.04 (0.46)	26; 51; 23	0.33	0.43	42.82	78 (0.2)
Bromo (Ar)	−0.03 (0.43)	21; 52; 27	0.8	0.89	78.82	425 (0.125)
Ethyne	−0.02 (0.28)	11; 76; 13	0.53	0.8	23.93	70 (0.5)
Methoxy	−0.01 (0.31)	14; 70; 16	−0.11	−0.05	29.9	1935 (0.125)
Ethoxy	0 (0.36)	18; 63; 19	0.51	0.56	43.96	147 (0.5)
Pyridyl	0 (0.58)	31; 43; 26	0.4	0.83	77.26	184 (0.5)
Fluoro (Ar)	0.01 (0.32)	17; 68; 15	0.12	0.11	17.86	2848 (0.2)
Chloro (Ar)	0.01 (0.4)	23; 54; 23	0.64	0.72	34.34	2438 (0.125)
Thiomethyl	0.03 (0.43)	33; 34; 33	0.57	0.67	46.05	48 (0.5)
Dimethylether	0.04 (0.39)	26; 56; 18	0.05	0.17	43.85	196 (0.2)
Methyl	0.06 (0.44)	26; 59; 15	0.34	0.47	13.93	6696 (0.00025)
Fluoro-methyl	0.07 (0.36)	17; 74; 9	0.15	1.33	31.84	23 (0.2)
iso-Propyl-oxy	0.09 (0.4)	33; 49; 18	0.55	0.54	57.61	51 (0.125)
Furan	0.12 (0.44)	24; 63; 13	0.89	1.23	66.59	46 (0.125)
Propyl	0.13 (0.53)	35; 41; 24	1.48	1.6	42.01	290 (0.0025)
Tetra-hydro-furan	0.14 (0.38)	28; 56; 16	0.24	1.14	69.7	25 (0.125)
t-Butyl	0.15 (0.51)	29; 58; 13	1.76	1.86	56.03	152 (0.0025)
Phenyl	0.15 (0.63)	40; 33; 27	1.76	1.9	76.07	588 (0.00025)
Ethyl	0.19 (0.57)	38; 45; 17	0.97	1.24	27.96	1083 (0.00025)
Thiophene	0.22 (0.59)	48; 27; 25	1.47	1.77	82.05	91 (0.0025)
Methyl-ester	0.23 (0.61)	41; 36; 23	0.13	0.45	57.86	183 (0.00025)
iso-Propyl	0.23 (0.58)	42; 40; 18	1.3	1.54	41.98	454 (0.00025)
i-Butyl	0.27 (0.63)	46; 35; 19	1.77	1.97	55.92	91 (0.00025)
Benzyl	0.34 (0.76)	54; 22; 24	2.01	2.33	89.99	212 (0.00025)
n-Butyl	0.4 (0.79)	57; 24; 19	2.06	2.36	56.15	131 (0.00025)
Cyclohexyl	0.41 (0.66)	51; 36; 13	2.34	2.58	80.73	69 (0.00025)
Cyclo-pentyl	0.45 (0.51)	66; 24; 10	1.75	1.73	67.78	29 (0.00025)

See Table 1 legend for specific details.

stated above, we believe caution is required so that such data are not over-interpreted.

2.2. Substitutions versus replacements

A limitation of the data presented in Tables 1–8 is that only transformations involving a hydrogen atom being substituted by a common chemical group have been considered, and not bio-isosteric replacements. However, if one makes the general assumption of additive SAR, the fundamental approximation of such methods, it might be possible to extrapolate the effect of replacing a substituent X with Y by subtracting Δact X–H from Δact Y–H. To assess this we have considered the substituents: methyl, chloro, phenyl and pyridyl, and investigated X–Y replacements using the P450 3A4 and solubility data sets. The results are summarized in Table 10 and vary according to the nature of X and Y. The conservative replacement of a methyl with a chloro can indeed be extrapolated using the results from Tables 1–8, the associated absolute error being <0.05 for both assays. When phenyl is replaced by pyridyl the extrapolation results are no longer consistent. The error associated with solubility is <0.1, however in the case of 3A4 inhibition the effect of replacing phenyl with a pyridyl is overestimated by +0.4 log units. A couple of considerations must be made here. Firstly, phenyl and pyridyl despite obvious similarities show some distinct property differences, in that pyridyl is more polar, can engage in H-bond interactions and chelate to haem groups. In addition, the cavity of 3A4 is the most accommodating of the P450 isoforms studied here, and is therefore most likely to exploit the differences between the two groups. It follows that trying to extrapolate the extent of the activity change is more challenging than when dealing with physico-chemical assays. Secondly, the *P* value associated to this transformation is only >0.2, therefore one cannot be confident at the 95% level that the mean differences are statistically significant. This is because of the high variability observed across the molecular pairs. We have further investigated this by considering the three different regioisomers of pyridyl. It is evident from this analysis that the average observed change for pyridyl is indeed dominated by the 2-pyridyl regioisomer alone (−0.02 and −0.15, respectively), as it accounts for 370 of the 488 molecular

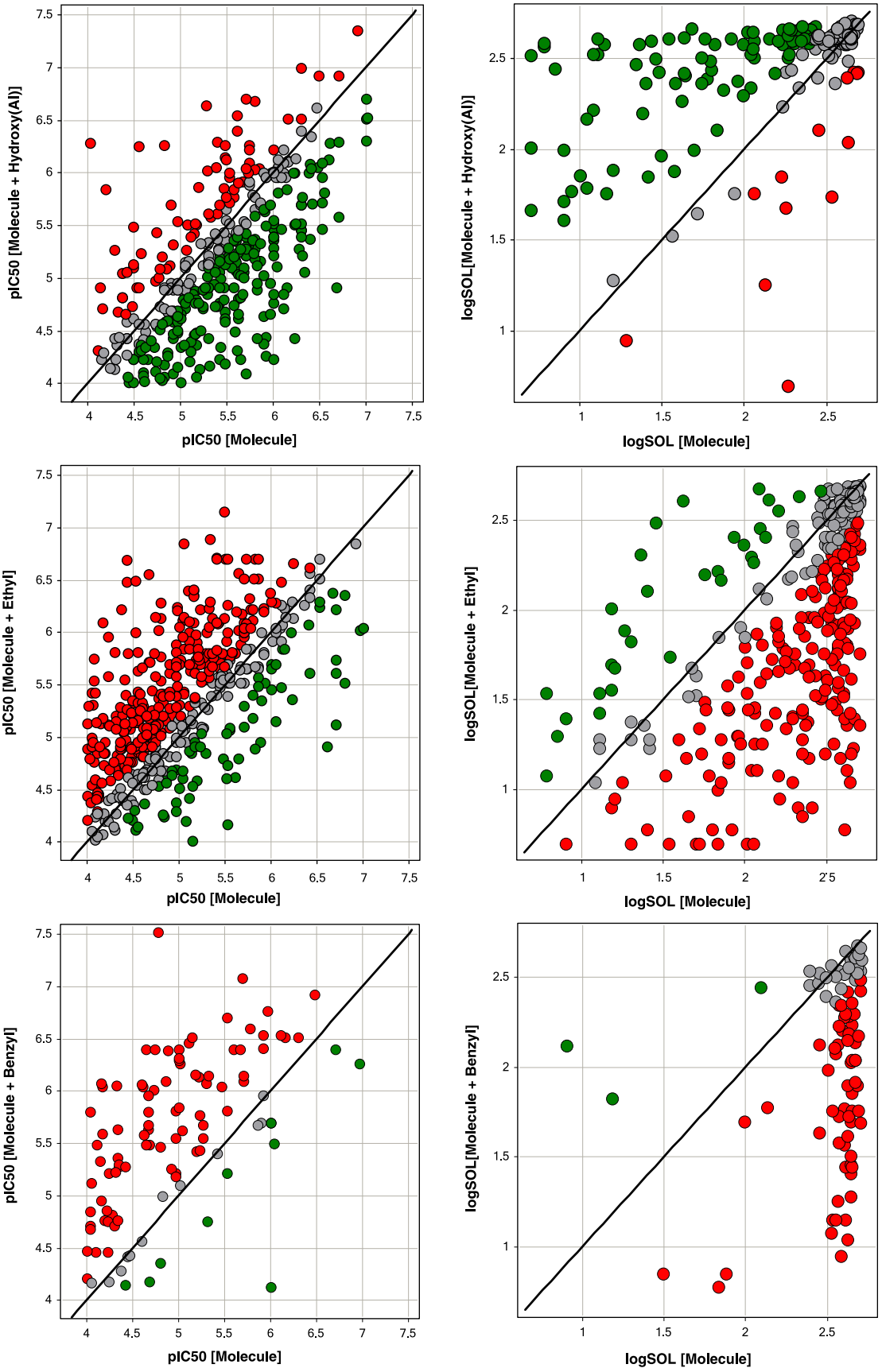


Figure 2. Plot of the experimental ADMET parameters (log Solubility (mM/L) and P450 3A4 pIC_{50}) for a given molecule and that with either an additional ethyl, aliphatic hydroxyl or benzyl substituent. Data coloured according to change in activity observed; pIC_{50} change +0.2 (red), +0.2 to −0.2 (grey) and −0.2 (green).

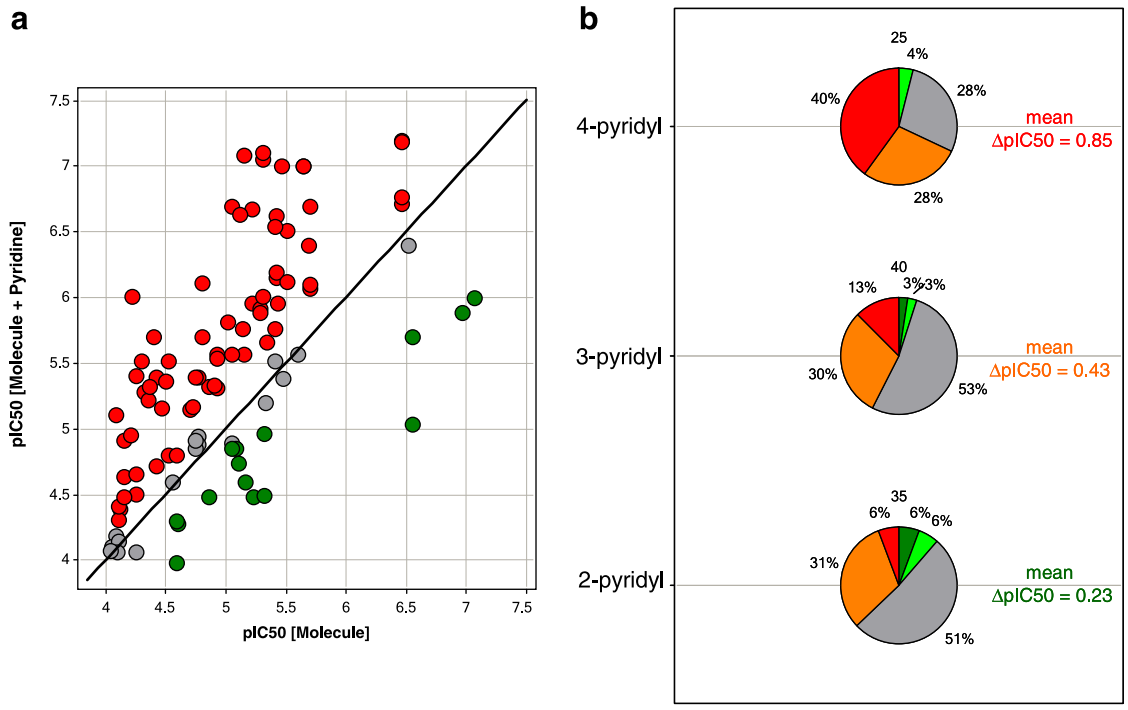


Figure 3. (a) Plot of the P450 2C9 pIC_{50} for a given molecule and that with an additional pyridyl substituent, coloured according to pIC_{50} change +0.2 (red), +0.2 to −0.2 (grey) and −0.2 (green). (b) The effect of replacing hydrogen with pyridyl on P450 2C9 inhibition (a). Data coloured according to ΔpIC_{50} >1.0 (red), 0.5 to 1.0 (orange), −0.5 to 0.5 (grey), −0.5 to −1.0 (light green) and <−1 (dark green).

pairs found. When the data are broken down by regioisomer, all the results are shown to be statistically significant (P >0.00025) and the differences between the three regioisomers are sizeable. 2-pyridyl causes a decrease of 3A4 inhibition with a concomitant drop in *c*log *D* (−0.98), whereas 4-pyridyl increases inhibition by ~1 log unit on average, presumably because of haem chelation, with 3-pyridyl occupying a position in between. A similar investigation using the solubility reveals equivalent trends, with the different pyridyl regioisomers showing statistically significant differences. Although less pronounced, the 4-pyridyl with the most accessible lone pair experiences the greatest solubility increase when substituted for phenyl.

These examples highlight the advantages of bio-isosteric replacements, in that the number of molecular pairs found when investigating substitutions may be far less than the number found when considering replacements. For instance, in the case of P450 3A4, the average effects of substituting hydrogen with phenyl and pyridyl have been derived from 257 and 57 pairs, respectively. On the other hand, 488 pairs have been found where a phenyl ring has been replaced by a pyridyl. This is an area in which we are currently undertaking additional research.

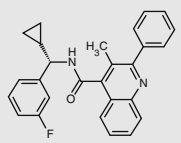
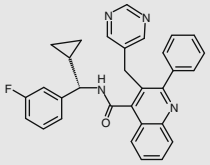
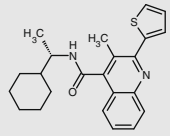
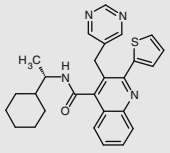
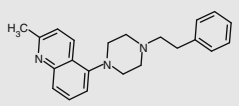
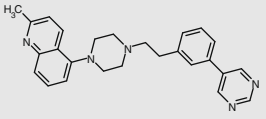
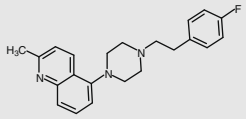
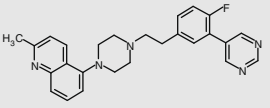
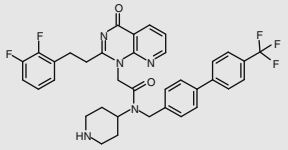
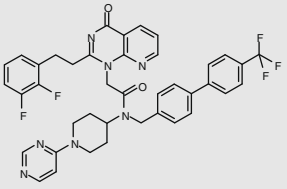
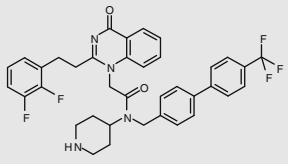
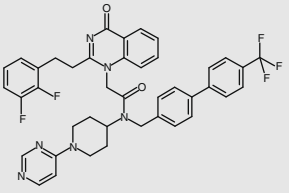
2.3. Relationship between ADMET assays and lipophilicity

The dominating effect of lipophilicity on in vitro inhibition assays can be illustrated clearly by plotting the average ΔpIC_{50}s reported in Tables 1–6 by the corresponding average Δclog *D*s at pH 7.4 (Fig. 4). Here the average ΔpIC_{50} of a given substituent is plotted against the average Δclog *D* for all the pairs identified to illustrate the formers' dependence on this key parameter. In addition, we colour each substituent according to the percentage change in activity: the cases which increase by +0.2 units are coloured red and those that decrease by −0.2, green. Apart from P450 1A2, and to a lesser extent 3A4 we observe a strong correlation between the average Δclog *D* of a substituent and the average ΔpIC_{50}, which is unsurprisingly reported elsewhere.[14] This is expected given that the larger the increase in lipophilicity, the greater the preference for the lipophilic environment of the P450 active site over aqueous solution.

P450 1A2 is a clear outlier, with the average ΔpIC50 displaying a poor correlation with the average Δclog *D*, Δclog *P* and ΔMWT and this is presumably a result of its small, restrictive active site. From Figure 4 it can be seen that larger substituents such as piperidinyl, *t*-butyl, morpholino, *i*-propyl and benzyl are identified as outliers, displaying large decreases in average ΔpIC_{50} when they replace hydrogen. It is also interesting to note that aromatic heterocycles such as pyridyl, pyrimidinyl and imidazolyl have a negligible effect on 1A2 inhibition, all displaying ΔpIC_{50}s no greater than +0.1 log units on average. This suggests that direct binding to the haem group is a rare occurrence due to the active site constraints. Generally speaking, substitution of any hydrogen atom for an alternative substituent is detrimental to 1A2 inhibition. In total, approximately 50% of the observed substitutions register a drop in ΔpIC_{50} >0.1 log unit, and a comparable increase of <20%.

P450 2C9 and 2C19 display high sequence homology (~91%)[23] so it is unsurprising that the results from replacing hydrogen with a range of substituents are similar. In both cases we observed a strong correlation between the average ΔpIC_{50} and Δclog *D*, however, certain outliers on both are apparent (Fig. 4). For example, pyridyl, pyrimidinyl and thiazolyl lie well above the ΔpIC_{50}/Δclog *D* regression line highlighting the uniqueness of these substituents. As mentioned earlier substitution of a hydrogen atom with a Lewis base capable of binding to the haem iron atom is a particular liability with P450s, especially with larger, more open, active sites such as 2C9 and 2C19. Three additional outliers in 2C9 are benzyl and the ionizable substituents carboxylic acid and piperidinyl. It is perhaps surprising to see carboxylic acid having such a negative effect on 2C9, compared to 2C19 as the latter protein is known to favour acids.[14] However, this analysis

Table 9
The effect of replacing hydrogen with pyrimidine on P450 2C9 inhibition

Structure one	Structure two	pIC_{50} one	pIC_{50} two	ΔpIC_{50}	Δclog *P*	Δclog *D*
 A1	 A2	5.1	5.0	−0.1	−0.9	−1.0
 B1	 B2	4.7	4.5	−0.3	−0.9	−0.5
 C1	 C2	4.4	4.7	0.4	−0.6	−0.2
 D1	 D2	4.5	5.0	0.5	−0.6	−1.0
 E1	 E2	4.7	7.0	2.3	0.3	2.4
 F1	 F2	4.8	7.1	2.3	0.3	2.4

Reported are the pIC_{50}, ΔpIC_{50}, Δclog *P* and Δclog *D* values.

does not take into account the absolute level of potency, rather it takes only the difference. In addition, only 23 pairs involving carboxylic acid in 2C19 were found, compared to 53 for 2C9, highlighting the greater propensity of the latter to tolerate acids, as expected.[24] The other interesting substituent is benzyl, which has a ΔpIC_{50} of +0.68 for 2C9, compared to +0.38 for phenyl alone. This is larger than the effect expected from their lipophilicity differences suggesting that the additional flexibility of the benzyl ring helps to strengthen interactions with aromatic residues within 2C9. The effect is also seen for 2C19 although the average ΔpIC_{50} values are somewhat lower at +0.47 and +0.38, respectively.

The average ΔpIC_{50} of substituents derived from the P450 2D6 data display the same strong relationship with the average Δclog *P*, as seen for 2C19 and 2C9. However, the structural alerts obtained in the latter cases are not observed with the former. Like the 1A2 isoform, Lewis bases do not pose as significant a risk due to the shape of the 2D6 cavity, minimizing the likelihood of chelation to the haem. Two of the most polar substituents, primary-sulfonamide ($R–S(=O)(=O)NH_2$) and carboxylic acid display the lowest

Table 10
The effect of replacing methyl with chloro, and phenyl with pyridyl on P450 3A4 inhibition and solubility

Transformation X–Y	ΔpIC_{50} 3A4 (SD)	*N* (>*P*)	Extrapolated value from (H–X) – (H–Y)[a]	Extrapolation error
Me ≫ Cl	0.05 (0.35)	385 (0.05)	0.10	0.05
Phenyl ≫ pyridyl	−0.02 (0.61)	488 (0.2)	0.40	0.42
Phenyl ≫ 2-pyridyl	−0.15 (0.48)	370 (0.00025)	–	–
Phenyl ≫ 3-pyridyl	0.26 (0.71)	95 (0.005)	–	–
Phenyl ≫ 4-pyridyl	1.00 (0.81)	23 (0.00025)	–	–
Transformation X–Y	ΔlogSol (SD)	*N* (>*P*)	Extrapolated value from (H–X) – (H–Y)[a]	Extrapolation error
Me ≫ Cl	−0.21 (0.40)	319 (0.00025)	−0.23	0.02
Phenyl ≫ pyridyl	0.42 (0.53)	634 (0.00025)	0.33	0.09
Phenyl ≫ 2-pyridyl	0.37(0.51)	436 (0.00025)	–	–
Phenyl ≫ 3-pyridyl	0.47(0.61)	130 (0.00025)	–	–
Phenyl ≫ 4-pyridyl	0.59 (0.53)	68 (0.00025)	–	–

In the case of pyridyl the 3 different regioisomers have also been investigated. The experimental values reported are the same as in Table 1.

[a] The extrapolated values are based on the values computed using Tables 5–8 by subtracting Δact X–H from Δact Y–H. The absolute error between observed and inferred value is also calculated.

average ΔpIC_{50} of all substituents. The latter is expected as 2D6 has a preference for bases, but the former is simply due to its polarity.

P450 3A4 is the most abundant P450[25] and has the largest active site cavity of the isoforms considered here. Thus, it is of little surprise to find that the majority of substituents here lead to an increase in pIC_{50} (Fig. 4 and Table 5). We find a weaker relationship between the average ΔpIC_{50} of a given substituent and the average $\Delta clog\,P$, but similar structural alerts are observed as found in 2C9 and 2C19. In addition, we only find four substituents that lead to an average ΔpIC_{50} <−0.2 log units. Indeed, 3A4 displays almost the opposite behaviour of 1A2, with substituents irrespective of their type, generally leading to increased 3A4 inhibition.

Analysis of the inhibition data compiled for the hERG potassium channel reveals a less complicated picture when compared to the five different P450s. The correlation between the mean ΔpIC_{50} of a given substituent and the average $\Delta clog\,D$ reveals almost a straight line (Fig. 4). Almost all of the pairs contain a basic centre so the ΔpIC_{50} of −0.07 for an aliphatic amine (dimethyl-amine) does not mean that a basic centre is beneficial to reduce hERG. Rather the introduction of an additional basic centre may not lead to increased hERG on average due to its impact on *c*log *D*. It is not clear to the authors why some subtle outliers are found: *n*-butyl with 62 diverse pairs, methyl-ester and acetyl.

The influence of lipophilicity on the two ADME parameters studied here can be appreciated by plotting the average ΔlogSOL and ΔlogPERM reported in Tables 7 and 8 by their corresponding average $\Delta clog\,D$ (Fig. 5). In both cases, the magnitude of the solubility or permeability change is directly proportional to the average change in *c*log *D* for a given substituent.

As the lipophilicity of a substituent decreases we observe the solubility to increase on average, in line with reports by others.[14,21] In Figure 5 it is observed that substituents such as carboxylic acid and aliphatic-amines lead to very desirable changes in solubility due to their dramatic impact on *c*log *D*. In contrast, substituents including phenyl and furan define the opposite extreme. A number of subtle outliers can be seen, including *i*-butyl and *n*-butyl, and although they are very detrimental to solubility, they appear to show a smaller change than would be expected based on their lipophilicity alone. This is presumably because of their steric bulk disrupting interactions in the crystal lattice thus benefiting the solubility.

Permeability is the only ADMET parameter studied here that shows an overall improvement with increasing *c*log *D* (Fig. 5). The carboxylic acid and aliphatic amine substituents were the most desirable replacements from a solubility point of view, but for permeability these represent the poorest choice.

3. Conclusions

A systematic analysis of ~500,000 ADMET datapoints, across eight in vitro assays has been undertaken to identify what is the impact of a range of common substituents on a range of ADMET parameters.

In this analysis we have identified all matched molecular pairs, differing by the replacement of a hydrogen atom with a list of predefined substituents. We have then generated a set of substituent tables which indicate the effects that may occur when a given substituent is introduced. These tables have been derived considering a wide range of chemotypes and paying attention to minimize issues about their generalizability; nonetheless, caution is required in extrapolating from this data set as these trends represent average changes only and will not hold up for each individual chemical series.

What is clear from the results is that there are no perfect substituents, that is, groups that can lead to a significant beneficial effect across all ADMET parameters. This is an obvious consequence of conflicting physico-chemical objectives, in that what is beneficial for say solubility is actually detrimental to permeability. The results are further confirmation that finding the right balance in properties is a huge challenge during lead optimization. It follows that developability properties should be investigated as early as possible to assess the probability of success of a given chemical series, and to prioritize different scaffolds.[13]

The data presented herein also indicate that the effect a substituent has on a given ADMET assay generally tracks with the change in *c*log *D*. Substituents with the highest *c*log *D* generally will have a greater detrimental effect than those that decrease *c*log *D*. The only case where this is not true is for permeability. In terms of the impact on lead optimization campaigns, from a consideration of all the SAR reported herein it is clear that optimizing chemotypes with insufficient target affinity and multiple ADMET issues is likely to prove extremely difficult. Programme teams that attempt to increase target affinity through the addition of lipophilic substituents will face challenges, as although this might prove beneficial for permeability, it would also be expected to lead to increased P450 inhibition and reduced solubility.

4. Computational procedures

4.1. Datasets and descriptors

The eight GSK ADMET datasets used in this study have been described in detail elsewhere.[14] The number of unique molecules measured in each ADMET assay ranges from ~10,000 to 50,000,

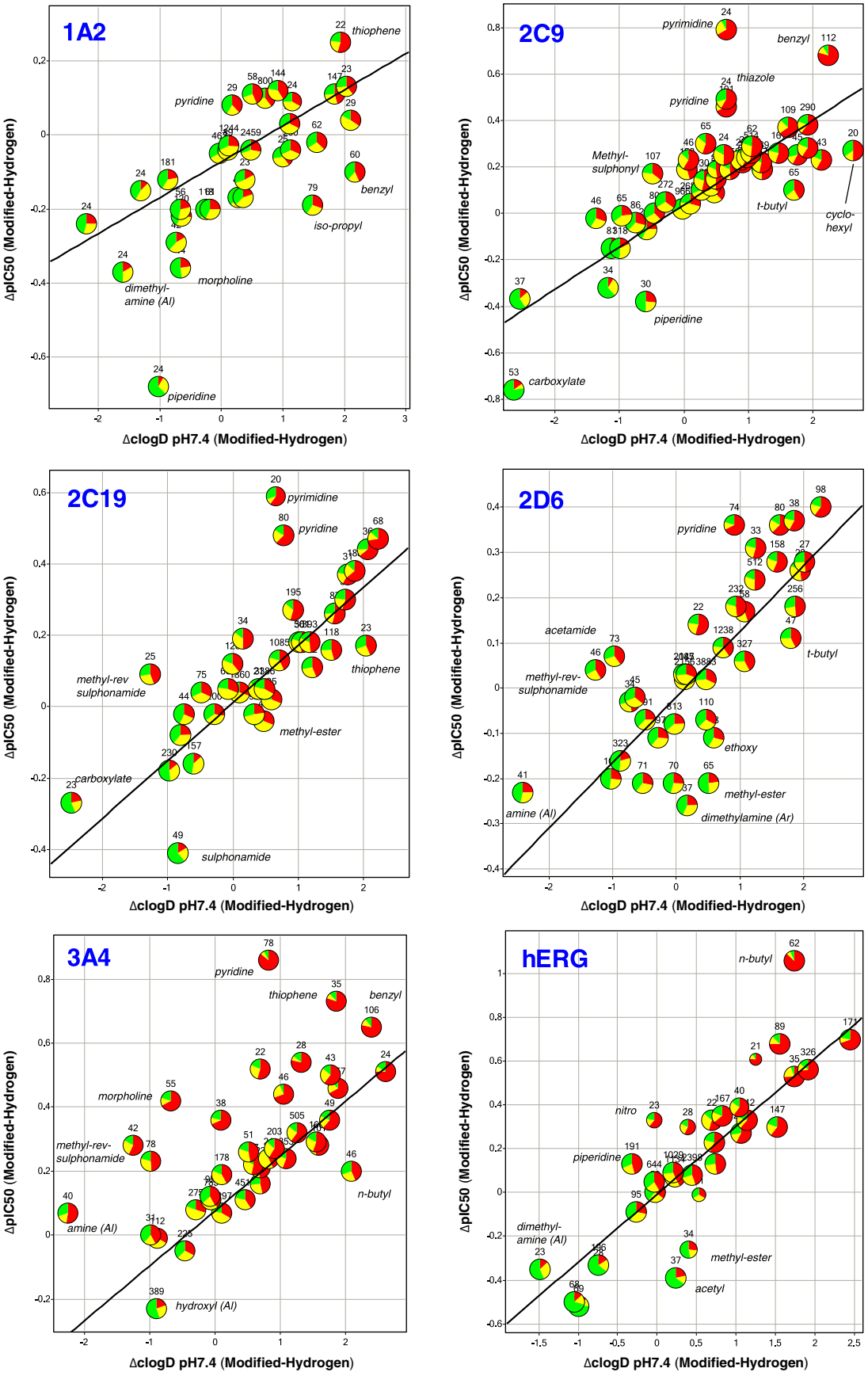


Figure 4. Plot of the average change in pIC_{50} versus the average change in clog *D* 7.4 when mutating a hydrogen atom to a given substituent. The percentage of cases that increase by 0.2 units are coloured red and those that decrease by 0.2, green. Substituent circles are sized according to pair count, the maximum size corresponding to >25 cases.

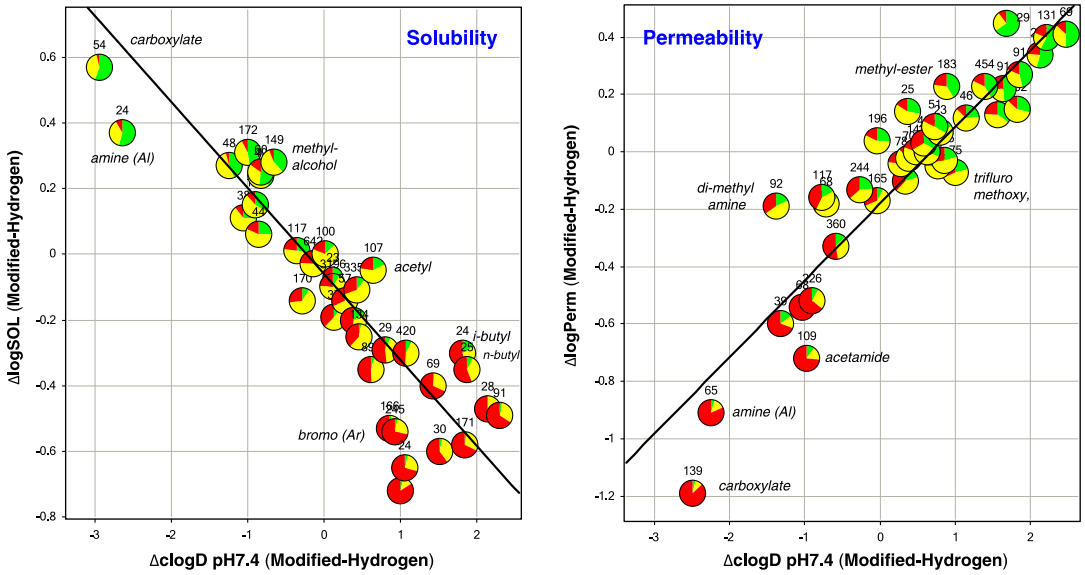


Figure 5. Plot of the average change in log (Solubility/mM/L)) and log (Permeability nm/s) versus the average change in *c*log *D* 7.4 following replacement of H with a particular substituent. For solubility the percentage of cases that increase by 0.2 units are coloured green and those that decrease by 0.2, red. The reverse scheme is used for permeability.

spanning a large range in response values and covering structurally diverse chemical space. For each molecule we have computed the Daylight *c*log *P*,[18] ACD *c*log *D* at pH 7.4[26] and the MWT.

4.2. Identification of matched molecular pairs

The *Find Magic* routine has been implemented as part of the *SAR-Toolkit*,[27] a GSK proprietary suite of applications for SAR data analysis, based on the Daylight Toolkit[28] and embedded within Spotfire DecisionSite.[29] *Find Magic* performs exhaustive pairwise comparisons and identifies matching pairs of molecules (M1 and M2) involving a single point change (group X in M1 replaced by group Y in M2, with the remaining structure being identical). Input molecules are represented by SMILES and transformations (X ≫ Y) by SMARTS. The routine is versatile, allowing the investigation of all possible structural transformations, where X can be hydrogen, a sidechain, a linker or a core, and Y can be a specific predefined substituent, or can be left unspecified (Y = any).

For computational speed, a pattern search is initially performed and only molecules containing groups X or Y are considered. Only suitable pairs of molecules are then compared, removing straight away those pairs that exceed thresholds on differences between atom and bond counts (thresholds are automatically defined based on the nature of X and Y). While similarity based on Daylight fingerprints can also be used as an initial filter to eliminate unsuitable pairs and speed up the process, they were not used in this investigation.[20]

In case Y ≠ any, X is removed from M1 and M1 associated with Y to give M1-X.Y, which is then converted to SMARTS for pattern searching of M2. If a match is found the pair is saved. Note that M1 may contain multiple entries of X, therefore this operation is repeated until a match with M2 is found (this is not necessary for the special case X = H). In case Y = any, only M1-X is used for pattern searching. If a match with M2 is found, M1 is subtracted from M2 to identify Y. The atom to which Y is attached to is also recorded and makes it possible to distinguish for instance between 2-pyridyl and 4-pyridyl, or between aromatic-amines and aliphatic-amines. In case X ≠ sidechain, an additional check is performed to verify that X and Y are similarly connected to the remainder of the molecule.

4.3. Analysis of molecular pairs

For this study we limit ourselves to transformations involving a hydrogen atom (X) being substituted by a predefined list of frequently used substituents (Y) (Supplementary data Table S1). We have not considered bio-isosteric replacements of substituents directly, instead investigating whether the results obtained for H to Y substitutions are additive, and so can be used to extrapolate X to Y changes.

For each of the structural transformations we calculate the average difference in the ADMET parameters for all pairs identified, making a statistical assessment of whether the differences are significant or not using a one-tailed Student's *t*-test in the same way as that used by Haubertin et al.[20] A 2-tailed *T*-test is typically used when one wishes to assess the significance of mean differences for cases where one does not know the direction of differences (i.e., test both tails of the distribution). However, in this case we know that by replacing Hydrogen with other substituents the mean differences will reflect the Δlog *P*, so a 1-tailed test can be considered more appropriate. We report the probability arising from the one-tailed Students *T*-test (*P*) along with the mean difference (i.e., the probability that the means are different). We consider a *P* value above the 95% confidence interval to be statistically significant (i.e., $P < 0.05$). One can calculate the two-tailed *P* value from the following equation ($P_{2\text{-tailed}} = P_{1\text{-tailed}} * 2$).

We assess the molecular diversity obtained for each substituent studied to ensure that the results can be generalized. We only report values for a substituent where the number of pairs is >20, in line with the approach taken by Leach et al.,[21] and where there are at least 5 unique chemotypes. Distinct chemotypes are identified based on similarity (sphere exclusion algorithm,[30] Daylight fingerprints[18] and a Tanimoto cut-off of 0.7). To aid interpretation we also report the percentage of cases where the activity difference is >0.2 and <0.2.

Acknowledgements

The authors would like to thank Drs. Anne Hersey, Ian Forbes and Graeme Archer for carefully reading early drafts of this manuscript and providing useful suggestions to improve it.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2009.07.002.

References and notes

1. Kola, I.; Landis, J. *Nat. Rev. Drug Disc.* **2004**, *3*, 711.
2. Li, D.; Kerns, E. H. *Curr. Opin. Drug Disc. Dev.* **2005**, *8*, 495.
3. Kubinyi, H. *Nat. Rev. Drug Disc.* **2003**, *2*, 665.
4. Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003**, *46*, 1250.
5. Proudfoot, J. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1087.
6. Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. *Angew. Chem., Int. Ed.* **1999**, *38*, 3743.
7. Leeson, P. D.; Davis, A. M.; Steele, J. *Drug Discovery Today* **2004**, *1*, 189.
8. Lajiness, M. S.; Vieth, M.; Erickson, J. *Curr. Opin. Drug Disc. Dev.* **2004**, *7*, 470.
9. Hann, M. M.; Oprea, T. I. *Curr. Opin. Chem. Biol.* **2004**, *8*, 255.
10. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **1997**, *23*, 3.
11. Hopkins, A. L.; Groom, C. R.; Alex, A. *DDT* **2004**, *9*, 430.
12. Abad-Zapatero, C.; Metz, J. T. *DDT* **2005**, *10*, 464.
13. Leeson, P.; Springthorpe, B. *Nat. Rev. Drug. Disc.* **2007**, *6*, 881.
14. Gleeson, M. P. *J. Med. Chem.* **2008**, *51*, 817.
15. Free, S. M., Jr.; Wilson, J. W. *J. Med. Chem.* **1964**, *7*, 395.
16. Lewell, X.; Judd, D. B.; Watson, S. P.; Hann, M. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511.
17. Japertas, P.; Didziapetris, R.; Petrauskas, A. *Quant. Struct.-Act. Relat.* **2002**, *21*, 23.
18. clog *P*. Chemical Information Systems Inc. Aliso Viejo, CA, http://www.daylight.com.
19. Sheridan, R. P. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103.
20. Haubertin, D. Y.; Bruneau, P. *J. Chem. Inf. Model.* **2007**, *47*, 1294.
21. Leach, A. G.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L.; MacFaul, P.; Wood, J. M.; Colclough, N.; Law, B. *J. Med. Chem.* **2006**, *49*, 6672.
22. Hajduk, P. J.; Sauer, D. R. *J. Med. Chem.* **2007**, *51*, 553.
23. Niwa, N.; Kageyama, A.; Kishimoto, K.; Yabusaki, Y.; Ishibashi, F.; Katagiri, M. *Drug Metab. Dispos.* **2002**, *30*, 931.
24. Lewis, D. F. *Biochem. Pharmacol.* **2000**, *60*, 293.
25. Shimada, T.; Yamazaki, H.; Mimura, M.; Inui, Y.; Guengerich, F. P. *J. Pharmacol. Exp. Ther.* **1994**, *270*, 414.
26. ACDlogD. Advanced Chemistry Development, Toronto, Ontario, Canada www.acdlabs.com.
27. Atkinson, F.; Bravi, G; The SAR Toolkit, UK-QSAR and ChemoInformatics Group meeting, October 2007: Abstract (http://www.iainm.demon.co.uk/ukqsar/meetings/2007-10-11.html#abstract_e013ca26c3a3c590184b907a3d2668c2) and presentation (http://www.documentarea.com/qsar/Francis_Atkinson07.pdf).
28. Daylight v4.92 Chemical Information Systems Inc. Aliso Viejo, CA, http://www.daylight.com.
29. Spotfire Decision Site 8.2.1, available from http://www.spotfire.com/.
30. Butina, D. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747.