Bioorganic & Medicinal Chemistry 18 (2010) 4405–4414

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

A statistical analysis of in vitro human microsomal metabolic stability of small phenyl group substituents, leading to improved design sets for parallel SAR exploration of a chemical series

Alexander G. Dossetter *

AstraZeneca PLC, Alderley Park, Mereside, Cheshire, SK10 4TG, United Kingdom

ARTICLE INFO

Article history:
Received 28 October 2009
Revised 1 March 2010
Accepted 25 April 2010
Available online 29 April 2010

Keywords:
ADMET
In vitro HLM stability
In vitro clearance
Matched molecular pairs

ABSTRACT

An analysis of in vitro human liver microsomal turnover assay results from a large dataset (~75 K) of experimental compounds tested is presented. Combined with an analysis of small (<6 Ha) substituents on known drugs and existing published results a new set of 29 substituents (*consensus*) is proposed to increase stability and probe SAR (an enhanced 'Topliss set'). In addition a different group of 28 substituents are identified as unlikely to change in vitro HLM stability, and a further set of compounds focuses on increasing HLM stability only.

1. Introduction

In 1972 Topliss published a seminal paper recommending a systematic stepwise method of building a structure–activity relationship for a chemical series.[1] At the time of that publication biological testing methods often had limited throughput so a focused low rate of synthesis was more in tune with the rate of testing. In recent years parallel synthesis, and more importantly purification, in addition to a larger pool of available reagents have vastly increased the number of compounds synthesised within projects. As a result of this, a 'blunderbuss' approach to SAR generation has been more common, with libraries of compounds synthesised, tested then analysed together in parallel, in line with efforts to streamline processes and efficiency. However the industry is under pressure to reduce costs, and timelines, whilst increasing the quality of compounds entering development.[2] Indeed efforts across several companies using the 'lean manufacturing' principles have led to closer attention being paid to the quality of compound design.[3–6] If a hit explosion campaign were to make 100 variants on a phenyl ring, test them in parallel, and find that the SAR was 'flat', that is, no gain in activity or any other property, then this would be wasteful. A leaner, more agile approach would initially assess SAR with a small subset to add 'high value' quickly or more importantly to stop work in an area.

To illustrate this, when the hit compound **1**, was varied to increase CCR5 antagonist activity (Fig. 1),[7] 89 compounds of the type **2** were synthesised. After testing and analysis no compound was more active than the initial 4-chlorobenzyl hit.[8]

A simple solution would be using the whole Topliss tree, which can be argued as very lean design, by using all 32 substituents in parallel synthesis. However within large pharmaceutical companies there exists a vast dataset of knowledge from the intervening 35 years of drug discovery, so it should be possible to define a new set of small group substituents that are more likely to yield 'drug like' properties. One such analysis has already been carried out by Hajduk and Sauer,[9] where they examined matched pairs of compounds for simple substituents anywhere within a molecule. From their analysis over a vast range of protein targets they produced two possible sets of 24 substituents for use in drug

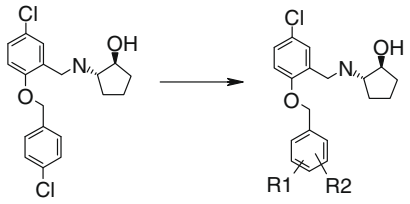


Figure 1. CCR5 antagonists library synthesis, where R1,2 = small groups (⩽5 Ha) or H.

* Tel.: +44 (0)1625517673; fax: +44 (0)1625513624.
E-mail addresses: al.dossetter@astrazeneca.com, al@dossetter.plus.com

0968-0896/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2010.04.077

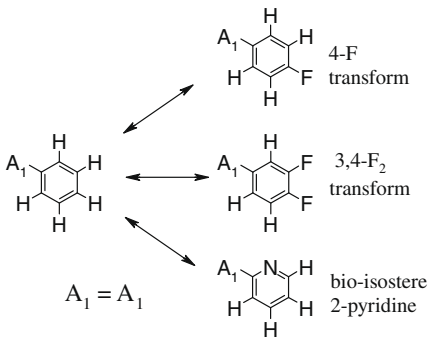
discovery; one set with an increased probability of gaining affinity and the other to examine diversity of substituents. In a similar vein a database of in vitro human liver microsomal stability data could be mined to examine which substituents gave a higher probability of increasing stability, and thus increased chances of yielding improved pharmacokinetic properties and low clinical dose. An analysis has already been carried out by Lewis and Cucurull-Sanchez, using the Pfizer corporate database of results, yielding probabilities of changing stabilities by >2-fold.[10]

We have attempted by looking further and in more detail using the AstraZeneca (PLC) database of human microsomal data and matched pair analysis, Hajduk and Sauer sets, and by looking at the frequency with which these substituents occur in known drugs as corroborating evidence to achieve the goal of an 'ideal' design set. Thus we propose a new refined set of 29 (21 from the original Topliss tree) substituents to examine substitution on a phenyl or potentially any aromatic ring.

2. Results and discussion

The AstraZeneca (PLC) global database of in vitro, in vivo biological, physical properties and pharmacokinetic test results contained ~75 K results for single compounds tested in an in vitro stability experiment using human liver microsomes.[11] These data were collated from 18 assays throughout the company over several years. To compare meaningful results and draw conclusions, matched pairs of compounds were found for a single change from pendant phenyl ring to the required combination of substituents for each of the groups from the Topliss tree; along with several di-substitutions, and other common phenyl bio-isosteres (Scheme 1); in total 73 ' transforms'.[12] The dataset used here predominates from single test results (96%), and further study found that the majority of matched pairs (89%) were tested in the same assay (i.e., same protocol, same laboratory site). Ideally to reduce error and increase robustness of the results and interpretation, all compounds would be tested multiple times in large batches, ensuring matched pairs are done within the same experiment. However, within each experiment standard compounds are run and the assay data have only been uploaded to the database where the standard result falls within a range of previous results (usually <2-fold). In a similar manner measured lipophilicities were extracted for these matched pairs, where available, (Log $D_{7.4}$, octanol) to derive a Δ Log $D_{7.4}$ to use in the analysis.

Results were standardised and expressed in units of μl/min/mg of microsomal protein. Each result was converted to a base Log_{10} and subtracted to look for fold changes [i.e., Δ Clint; improvement = lower HLM Clint +ve value or detrimental = higher HLM Clint −ve value] and plotting all the results for a group of matched pairs as a distribution and conducting matched pair analysis



Scheme 1. Illustrated matched pair searches performed. A1 was the rest of the compound structure matching across transform.

(Student *t*-test—*p*-values reported). In this manner the distributions were checked for normality and then compared by the mean, standard deviation (SD) and standard error (SE). The dynamic range of the HLM assay data was 2.5 Log units (2.0–600 μl/min/mg), and 12% of the dataset were results where the value was out-of-range of the assay condition, that is, either below limit of detection to observe any change in parent compound (very stable) or completely consumed before the first time point measured (highly unstable). Both analyses of the results were conducted including and excluding the out-of-range values, with little change in means or SD, for large groups of matched pairs. The information recorded in out-of-range results was useful when scoring the dataset for frequency (likelihood) or probability of making a significant change in HLM stability. We chose to set a 'cut-off' of 0.5 Log_{10} unit change in stability and recorded the frequency [$F(0.5)$] above and below [$F(-0.5)$] this in Table 1 for each transform (see Section 5). Hence the statistical results tabulated in Table 1 are for the distribution of results excluding the out-of-range values, but the probability scores are including the out-of-range results.

The MDDR3D database of known drugs was sub-structure searched for substituted aromatic rings.[13] At the time, the database contained 1820 drug structures that had reached the market, and 1027 of these had an embedded phenyl ring structure. Of these 277 had a pendant phenyl ring, and of the remaining, 390 had a single point of attachment to 'the rest of the drug molecule' or scaffold. The remaining phenyls were polysubstituted or bicyclic. The exact search of the group transform (e.g., 4-fluoro-phenyl) was performed and those hits checked by eye. The functional group was then tested in any possible substituent pattern, and then any possible type of aromatic ring. Thus of the 70 groups/transforms (excluding isosteres) studied these accounted for 286 of the 390 compounds identified above. Sheridan looked for known drugs of the same biological activity to form pairs of compounds searching for bio-isosteres replacement groups.[14] In this case we have looked for only the occurrence of the small groups under study here, in the hope that the frequency bears some relationship to those that score well in our analysis, and ensure our designs set 'capture' them.

Overall the matched pairs plotted together as a single distribution Δ Log_{10} HLM Clint showed that, for the majority of time, the change in substituent did not affect a large change in metabolic stability, that is, more than a Log unit change (Fig. 2). The mean of the distribution was very close to zero with a SD of 0.45 (whole dataset). In fact only 138 matched pairs were shown to have a Log unit or more improvement in HLM stability (2.6% of the 5321 matched pair set), and equally 124 matched pairs had a detrimental effect on stability (2.3%). This clearly provides our first conclusion, that the chance of increasing metabolic stability for a lead series by more than a Log unit is fairly low, but not impossible, with the simple groups or isosteres studies here. This also re-enforces the need for metabolite identification in the cases where turnover is high to pin-point the site to be modified, but this is far from guaranteed. It is noteworthy to point out that the majority of the HLM data for the above 262 matched pairs at the extremes of the distribution were $n = 1$ results, and as such some maybe erroneous. Given the above analysis, cumulative probabilities for each substituent transform achieving greater than 0.5 Log unit change (3.2-fold) in HLM stability were calculated as this was greater than one standard deviation, and higher than the HLM assay protocol error. In the lead optimisation phase of discover a 0.5 Log unit change maybe all that is required for optimal stability of a new candidate. Despite this lower threshold, $F(0.5) = 0.079$ (i.e., 7.9%) and $F(-0.5) = 0.059$, we again see a low probability of increasing stability with small substituent such those examined here.

Table 1 details all of the data for each of the transforms studied. Lewis et al. have published the probabilities based on Pfizer global

Table 1

No.	Group[a]	Top[b]	H[c]	C[d]	U[e]	K[f]	Kn (any)[g]	N[h]	T[i]	Mean[j]	SD[k]	SE[l]	S2[m]	Δ Log *D*[n]	Log *D* dev[o]	*F* (0.5)[p]	S3[q]	*F*(−0.5)[r]	S4[s]
1	2-CH_3	1			*	7	48(127)	99	0.738	−0.12	0.43	0.04	**	−0.05	0.51	0.060		0.138	***
2	3-CH_3	1			*	5		80	0.736	−0.28	0.40	0.04	***	−0.27	0.44	0.000	**	0.186	***
3	4-CH_3	1			*	17		181	0.716	−0.24	0.45	0.03	***	−0.41	0.52	0.029	**	0.239	***
4	4-CH_2CH_3		1			2	3(10)	19	0.695	−0.39	0.60	0.14	*	−0.46	0.66	0.050		0.450	***
5	4-$CH(CH_3)_2$	1				0	5(13)	19	0.649	−0.47	0.76	0.18	*	−0.97	0.67	0.143		0.524	***
6	4-$C(CH_3)_3$		1			4	6(10)	23	0.709	−0.19	0.50	0.10	*	−1.25	0.86			0.240	***
7	3,4-$(CH_3)_2$	1		**	*	0	27(10)	11	0.688	0.03	0.30	0.09		−0.01	1.22	0.000		0.077	
8	4-Cyclopropyl		1	**		0	0(8)	20	0.674	0.08	0.67	0.15		−0.93	0.37	0.182		0.152	*
9	3-CF_3	1	1	**		6	25	94	0.699	−0.13	0.52	0.05	*	−0.90	0.46	0.087		0.204	***
10	4-CF_3	1		***		3		109	0.694	0.04	0.56	0.05		−0.79	0.56	0.176	**	0.137	***
11	2-CN				*	0	8(10)	58	0.688	−0.18	0.44	0.06	*	−0.01	1.14	0.048		0.222	***
12	3-CN				*	0		100	0.704	−0.01	0.46	0.05		0.16	0.63	0.125		0.134	***
13	4-CN	1		**		3		168	0.687	0.25	0.49	0.04	***	0.28	0.52	0.193	***	0.004	
14	4-$C(O)CH_3$	1			*	1	3(4)	16	0.709	−0.18	0.32	0.08	*	0.42	0.36	0.000		0.100	
15	3-$C(O)NH_2$			*		0	4(12)	14	0.670	0.09	0.59	0.16		1.29	0.59	0.267		0.200	**
16	4-$C(O)NH_2$	1		**		0		17	0.702	0.49	0.39	0.10	***	0.57	0.95	0.556	***	0.000	
17	4-CO_2H			*		5	26(33)	27	0.688	0.77	0.51	0.10	***	2.64	0.83	0.655	***	0.000	
18	2-F		1		*	8	74	169	0.699	−0.10	0.32	0.02	***	−0.12	0.44	0.020	**	0.083	**
19	3-F				*	1		234	0.682	−0.04	0.38	0.02		−0.19	0.43	0.049		0.052	
20	4-F	1		***	*	35		497	0.695	0.06	0.36	0.02	***	−0.18	0.44	0.086		0.035	
21	2-Cl	1	1	**	*	15	147	122	0.705	−0.19	0.49	0.04	***	−0.25	0.51	0.063		0.197	***
22	3-Cl	1		**	*	4		168	0.712	−0.14	0.45	0.03	***	−0.53	0.60	0.073		0.136	***
23	4-Cl	1		***	*	45		337	0.705	0.01	0.46	0.02		−0.57	0.51	0.079		0.097	***
24	2-Br		1			1	18	13	0.702	−0.12	0.60	0.17		−0.21	0.34	0.067		0.133	*
25	3-Br	1		**	*	0		28	0.677	−0.28	0.30	0.06	***	−0.52	0.41	0.000		0.226	***
26	4-Br	1		***		3		41	0.714	0.02	0.50	0.08		−0.84	0.40	0.152		0.196	***
27	3-I	1				0	18	4	0.673	−0.26	0.56	0.28		−1.26		0.000		0.500	***
28	4-I	1				0		9	0.697	−0.02	0.74	0.25		−1.19	0.34	0.222		0.222	**
29	2-Cl,4-Cl	1	1	**		15	36	42	0.689	−0.14	0.59	0.09		−1.43	0.24	0.037		0.185	***
30	3-Cl,4-Cl	1			*	1		104	0.698	−0.20	0.42	0.04	***	−1.21	0.74	0.035		0.181	***
31	2-Cl,4-F					0		18	0.639	−0.08	0.53	0.13		−0.98	0.42	0.130		0.261	***
32	3-Cl,5-Cl			*		0		27	0.701	0.09	0.62	0.12		−0.54	1.34	0.200	*	0.167	**
33	2-Cl,6-Cl					9		11	0.707	−0.28	0.52	0.16		−0.89	0.62	0.077		0.308	***
35	2-F,3-F				*	0	17	44	0.664	−0.11	0.41	0.06	*	−0.20	0.40	0.089		0.133	**
36	3-F,4-F				*	0		119	0.683	−0.01	0.39	0.04		−0.39	0.67	0.113		0.038	
37	3-F,5-F				*	0		48	0.660	−0.04	0.31	0.04		−0.15	0.78	0.020		0.061	
38	2-F,6-F				*	1		29	0.672	−0.14	0.35	0.06	*	−0.03	0.44	0.003		0.094	
39	3-CF_3,4-Cl	1		**		1	—	6	0.649	0.37	0.85	0.35		−1.60	0.49	0.250		0.125	
40	2-OCH_3	1	1	**	*	9	75	151	0.688	−0.16	0.38	0.03	***	0.18	0.49	0.041		0.111	***
41	3-OCH_3				*	2		183	0.687	−0.23	0.38	0.03	***	0.03	0.38	0.010	**	0.185	***
42	4-OCH_3	1		**	*	9		299	0.694	−0.10	0.47	0.03	***	−0.03	0.46	0.073		0.132	***
43	2,3-$(OCH_3)_2$		1	**		0	43	14	0.646	0.05	0.64	0.17		0.22	0.23	0.250	*	0.125	
44	2,4-$(OCH_3)_2$			*		0		27	0.650	0.21	0.61	0.12	*	0.07	0.42	0.330	***	0.185	***
45	3,4-$(OCH_3)_2$			*		12		59	0.688	0.09	0.55	0.07		0.60	0.36	0.143		0.110	**
46	2-OCF_3		1		*	0	1	14	0.617	−0.22	0.40	0.11	*	2.00	.	0.000		0.200	**
47	3-OCF_3				*	0		20	0.621	−0.20	0.47	0.10	*	−0.79	0.54	0.000		0.227	***
48	4-OCF_3			**		0		40	0.655	0.16	0.57	0.09	*	−0.76	1.30	0.244	**	0.111	*
49	2-NH_2				*	0	27 (88)	4	0.680	−0.02	0.42	0.21		0.50		0.000		0.200	*
50	3-NH_2			*		0		11	0.678	0.44	0.70	0.21	*	0.90	0.08	0.364	**	0.091	
51	4-NH_2	1		**		11		10	0.713	0.17	0.48	0.15		0.96	0.40	0.273	*	0.091	
52	2-$N(CH_3)_2$		1			0	5	2	0.774	−0.69	0.75	0.53		−0.38	0.20	0.000		0.500	**
53	3-$N(CH_3)_2$	1		**		0		10	0.653	−0.28	0.45	0.14	*	−0.35	0.25	0.000		0.200	**
54	4-$N(CH_3)_2$	1				4		30	0.679	−0.14	0.64	0.12		−0.01	0.35	0.133		0.267	***
55	3-$NHSO_2CH_3$			*		0	6	9	0.551	0.42	0.97	0.32		0.73	0.21	0.300	*	0.100	
56	4-$NHSO_2CH_3$			*		3		9	0.586	0.68	0.58	0.19	*	1.06	0.18	0.556	***	0.000	
57	2-NO_2					3	35	13	0.674	−0.31	0.48	0.13	*	−0.03	0.62	0.071		0.357	***
58	3-NO_2	1		**		11		11	0.647	0.10	0.73	0.22		0.10	0.38	0.077		0.015	*
59	4-NO_2	1		**		2		27	0.668	0.12	0.51	0.10		0.04	0.47	0.121		0.061	
60	2-SCH_3		1			0	1	5	0.591	−0.72	0.30	0.13	*	−0.19	0.51	0.000		0.600	***
61	3-SCH_3					0		9	0.690	−0.55	0.24	0.08	***	−0.55	0.80	0.000		0.556	***
62	4-SCH_3					1		17	0.691	−0.36	0.55	0.13	*	−0.76	0.13	0.056		0.389	***
63	3-SO_2CH_3		1	**	*	0	3	32	0.644	0.04	0.38	0.07		1.18	0.77	0.061		0.152	**
64	4-SO_2CH_3	1		***		2		77	0.643	0.30	0.59	0.07	***	1.12	0.60	0.329	***	0.071	
65	3-SO_2NH_2		1	**		0	27	2	0.652	0.56	1.03	0.73		1.09	0.14	0.333		0.000	
66	4-SO_2NH_2	1		***		2		18	0.658	0.41	0.45	0.11	**	1.56	0.36	0.579	***	0.000	
67	4-Ph		1	**		4	N/A	23	0.707	0.35	0.72	0.15	*	−1.46	0.79	0.280	**	0.125	*
68	2-Morpholine		1			0	5	8	0.625	−0.60	0.51	0.18	*	−0.33	0.56	0.000		0.556	***
69	3-Morpholine					0		17	0.547	−0.23	0.53	0.13	*	0.24	0.17	0.059		0.235	***
70	4-Morpholine			**		0		22	0.609	0.05	0.61	0.13	**	0.33	0.39	0.273	**	0.136	*

(*continued on next page*)

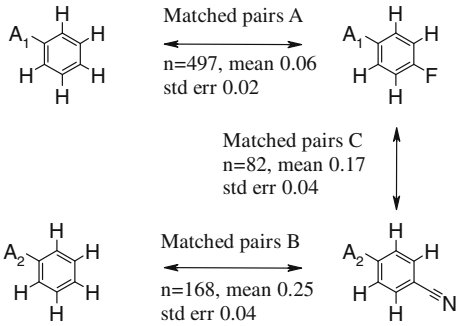
Table 1 (*continued*)

No.	Group[a]	Top[b]	H[c]	C[d]	U[e]	K[f]	Kn (any)[g]	N[h]	T[i]	Mean[j]	SD[k]	SE[l]	S2[m]	Δ Log *D*[n]	Log *D* dev[o]	*F* (0.5)[p]	S3[q]	*F*(−0.5)[r]	S4[s]
71	2-Pyridine			*		18		205	0.666	−0.05	0.45	0.03		0.78	0.62	0.107		0.068	*
72	3-Pyridine			*		17		200	0.687	−0.01	0.49	0.03		0.90	0.58	0.135	*	0.098	**
73	4-Pyridine					5		145	0.685	−0.08	0.60	0.05		0.79	0.75	0.185	**	0.142	***

[a] Modification from (5H)-phenyl.
[b] Member of the top 32 modifications described by Topliss—matched pairs were not found for 4-NO_3–3-CH_3 and 3-CH_3, 4-$(N(CH_3)_2$ groupings.
[c] Identified by Hajduk and Sauer as a positive group for increasing binding affinity.
[d] Combined rating = *** within Topliss set, identified by Hajduk and Sauer and a positive or neutral effect on HLM stability; ** two out of three of the criteria; * positive effect on HLM stability only.
[e] U—unlikely to change HLM stability, set—selected by mean between −0.20 and 0.06, SD ⩽ 0.45.
[f] Number of known drugs in MDDR3 db with this exact matching group on a phenyl ring.
[g] Number of known drugs with that functional group and any other substitution on a phenyl ring (no. of known drugs with this functional group on any aromatic ring).
[h] Number of matched pairs found with HLM data.
[i] Average Tanimoto distance across each set of matched pairs where distance = 0 was an exact match.
[j] Mean (Δ(Log10_mean clints for sub_phenyl)).
[k] SD = Std Dev(Δ(Log10_mean clints for sub_phenyl)).
[l] SE = Std Err(Δ(Log10_mean clints for sub_phenyl)).
[m] *p*-Value significance *** <0.001, ** <0.01, * <0.5.
[n] Mean(Δ(Measured_Log *D*_pH7.4)).
[o] SD(Δ(Measured_Log *D*_pH7.4)).
[p] Cumulative frequency of achieving at least a 0.5 Log unit improvement in HLM stability.
[q] Significance rating *p*-value for 2 × 2 contingency testing against the 4-fluoro dataset for *F*(0.5).
[r] Cumulative frequency of at least a 0.5 Log unit decrease in HLM stability.
[s] Significance rating *p*-value for 2 × 2 contingency testing against the 4-fluoro dataset for *F*(−0.5).

data for several of the transforms here using a cut-off of >2-fold change in stability (however without statistical data). For the most part we have found similar frequencies of changing stability by 0.5 Log unit [*F*(0.5)] for many transforms. For example the common method of increasing stability of a phenyl compound by the transform 4-fluoro was found to have 497 matched pairs (entry 20) with a mean of 0.06, SD 0.36, SE 0.02 (p <0.001).[15,16] The frequencies or tails of the distribution [*F*(0.5) = 0.086, *F*(−0.5) = 0.035], were surprisingly modest. Compare this to the transform 4-cyano (entry 13) with a positive shifted mean change of 0.25 (SD 0.49, SE 0.04, p <0.001), and *F*(0.5) = 0.193. It is clear from this 4-cyano provides a better prospect for increasing stability than 4-fluoro. However it was surprising that the means from these distributions was so close to zero. It could be incorrect to directly compare these two distributions as they do not necessarily contain the same core structures (Scheme 2—A1 may not equal A2). However from within the dataset we extracted matched pairs for the transform (4-F-Ph to 4-CN-Ph) and found n = 82, and a mean of 0.17 Δ Log_{10} HLM (SE 0.04, p <0.001), thus confirming that within SE the comparison can be made between these transform matched pairs.

With caution we can go further and consider the position and number of fluorine atoms on the phenyl ring by comparison of their distribution (Fig. 3). Comparing 4-fluoro substitution with 2 and 3, we can see distinct differences in the substitution pattern



Scheme 2. Illustration of how the matched pair analysis was performed for the paper. For matched pairs A: A1 = A1 and matched pairs B A2 = A2, however A1 and A2 may overlap for some pairs. For matched pairs C A1 = A2.

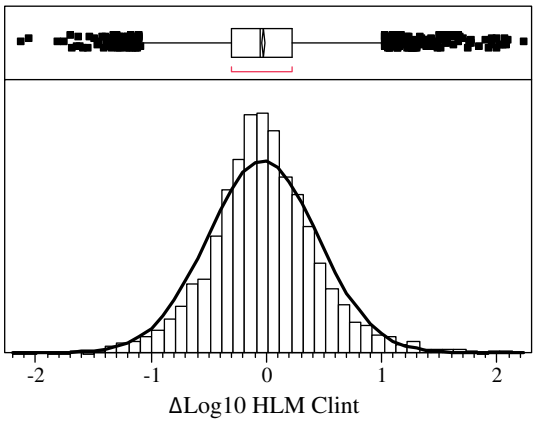


Figure 2. Distribution of the entire dataset of matched pairs (n = 5321). Mean Δ Log_{10} HLM Clint = −0.03, SD 0.48, SE 0.006.

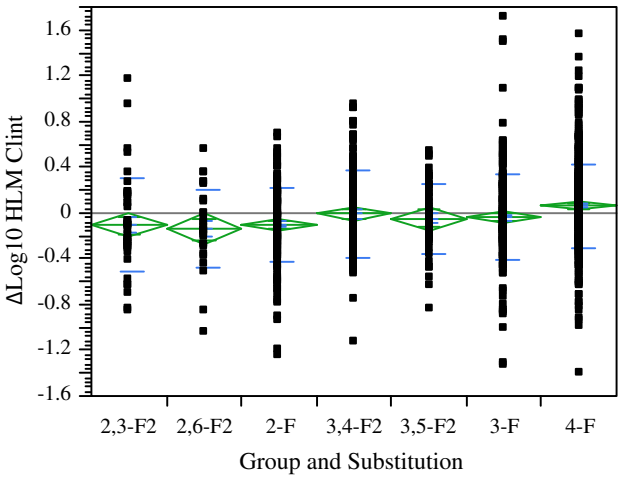


Figure 3. Comparison of means and distributions for all fluorine substitutions. Mean Anova diamonds are shown and blue markers as the SD for each spread of datapoints.

(confirmed using the Tukey and Kramer[53] method of comparing means), and as expected the 4-fluoro transform tended to increase stability to a larger extent. Di-fluoro substitution appeared no better than 4-fluoro having lower means and no higher frequency of matched pairs with an increase in stability (compare entry 20 with 35–38). We might have expected that two fluorine atoms would further reduce the electron density on a phenyl ring, and thus would be considered as less susceptible to oxidation. Without detailed metabolite identification we cannot determine if the site of primary metabolism on phenyl has indeed been blocked, or may have moved to another position within the molecule, as with any of the transforms studied here. This illustrates the difficulty of analysing these in vitro data, even though it is more useful and closer to the real clinical situation. The mechanisms for P450 ion haem oxidation have been extensively studied and reviewed elsewhere, and in light of this attempt to offer a practical route to compound optimisation, will not be discussed further.[17–25]

Analysis of all the transforms with data for 2, 3 and 4 positions found the same pattern with a higher likelihood of stability by substitution at the 4-position (Br, Cl, CN, $N(CH_3)_2$, OCF_3, NO_2 and SCH_3) with one notable and statistically tested exception, 2-CH_3 (entry 1 compared to 2 and 3). Whilst overall the 2-CH_3 group had a negative mean (−0.12, SD 0.40), compared to the other methyl substitutions, the distribution was shifted in favour of stability. It was unclear why similar sized groups (Cl and CN) do not have this effect.

For visual interpretations, the data for mean Δ Log_{10} HLM Clint has been plotted against SE (Chart 1) and SD (Chart 2) respectively. This was merely undertaken to spread the data out with the ordinate representing positive or negative effects, and the abscissa representing the degree of reliability by width of the distribution. This illustrates that many of the transforms studied have means close to zero with some degree of reliability. Thus this cohort of simple transforms is the most powerful knowledge we can derive from this analysis in that they will *probably not affect* the HLM stability. These 28 groups are highlighted (single *) in Table 1 under column U, and were selected with a mean between −0.20 and 0.06, SD ⩽ 0.45 [three of the transforms had low matched pair numbers but high diversity, so were included (3,4-$(CH_3)_2$; 2-Cl,4-F; 2-NH_2].

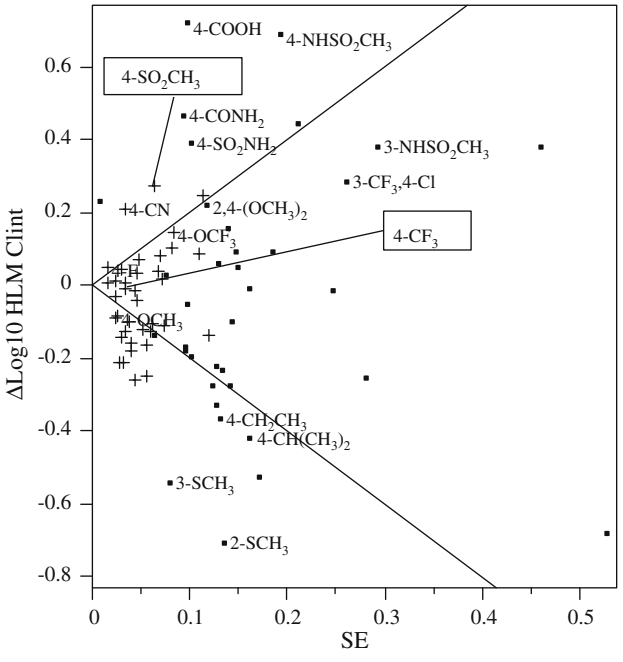


Chart 1. Spread plot of distribution mean Δ Log_{10} HLM against SE. Fitted lines (dotted) are a slope of 2 and -2 with origin of zero, that is, everything above, below or on the lines was a significant change in stability. + Marker $n \geqslant 30$ matched pairs.

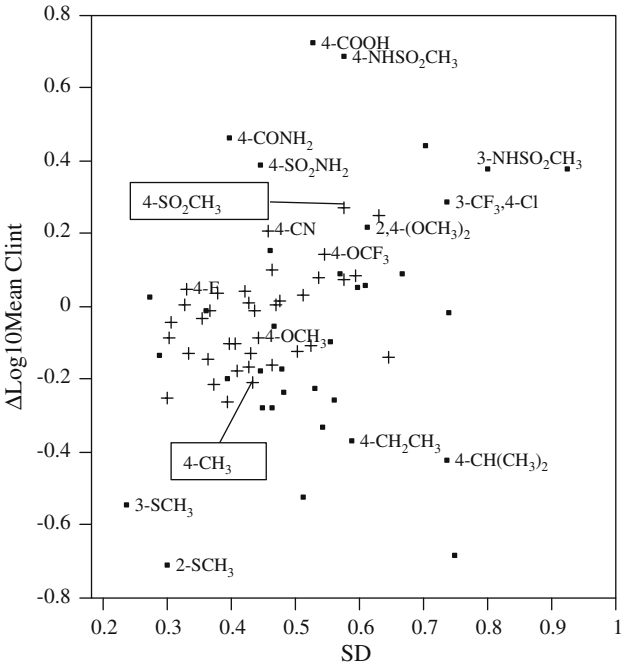


Chart 2. Spread plot of distribution mean Δ Log_{10} HLM against SD. + Marker $n \geqslant 30$ matched pairs.

We see that many of these transforms (CH_3, F, Cl, OCH_3, di-substitution) occur frequently in known drugs (qualitatively) so we can conclude that the scaffold or *core* of the compound is probably the governing factor, and these groups were probably added for binding affinity and/or selectivity.

Examining entries 3 through 7 (Table 1) we can see the effects of 4-substituted alkyl groups (these have been highlighted in Charts 1 and 2). As expected, instability rose (distributions shifted negatively) with the change from 4-CH_3, 4-CH_2CH_3 to 4-$CH(CH_3)_2$ which was consistent with an oxidation mechanism on the carbon attached to the phenyl ring, and increased ability to stabilise a reactive intermediate.[26] However it was interesting to note that there was little difference between 4-$C(CH_3)_3$ and 4-CH_3, and the former has been cited as a group prone to higher turnover.[27] Of particular note was 3,4-$(CH_3)_2$ (entry 7, mean 0.03, SD 0.27), where perhaps the presence of two methyl groups provided a 'co-steric block' to each other. In comparison 4-cyclopropyl (entry 8) was similar in neutral effect, which was again consistent with reduced ability to stabilise a reactive intermediate during the oxidative event.[28,29] In contrast the 4-CF_3 group was more robust with slightly positive mean but higher $F(0.5)$ of 0.176. It would therefore be prudent for lean design covering lipophilic substitution to proceed straight to 3,4-$(CH_3)_2$, 4-cyclopropyl and 4-CF_3.

Comparing the distributions for 4-OCH_3 (entry 42) versus 4-OCF_3 (entry 48) we can see that 4-OCH_3 had a tendency to decrease stability, but 4-OCF_3 by contrast was statistically biased towards increasing stability (mean 0.16, SD 0.57, $F(0.5) = 0.24$). Whilst the discussion above has focussed on protection of the phenyl ring, in this case addition of a methoxy presents the opportunity for de-methylation, which it can be argued that OCF_3 does not, and was consistent with an example from the literature.[30] The largest and most significantly shifted distribution was for the 4-CO_2H transform with a mean Δ Log_{10} HLM Clint 0.70 (SD 0.53, $n = 29$). This was not surprising given that the group would be ionised within the assay conditions, and thus may have strong repulsion within the P450 enzyme, as an end point of oxidation. We suggest it is treated separately from other transforms within this table as the medicinal chemist must consider the implications of

making the compound acidic depending on other functionality within the molecule. Morpholine and 4-phenyl matched pairs were added to the dataset to represent large volume groups. 4-Morpholine (entry 70) transform gave a mean around zero and broader distribution (SD 0.61) such that $F(0.5)$ was found to be 0.27, with $n = 22$ matched pairs. More dramatically the 4-Ph transform (entry 67) had a large positive mean (0.35) and broad distribution such that $F(0.5) = 0.28$. Clearly these two large groups have bigger impact and lend weight to the argument that overall molecular recognition is key.

The transforms to either 2-SCH_3, 3-SCH_3 and 4-SCH_3 groups had large negative means and SD (entries 60–62), clearly showing they are reliably detrimental ($F(-0.5)$ 0.60, 0.55 and 0.38, respectively), presumably due to their own oxidation or electronic activation of the phenyl ring. However the data must be treated with care as there were less than 30 matched pairs in all cases even though the conclusion is reasonable. In total contrast the 4-SO_2CH_3 group (entry 64) had a good population of matched pairs ($n = 85$), positive mean (0.30, $p < 0.0001$) and $F(0.5) = 0.329$. We can conclude that the 4-SO_2CH_3 increases stability with a reasonable degree of probability, which can be explained by either the large size of the group, electron withdrawing nature or simply by reducing the overall lipophilicity of the compound. Finally, as a comparison to all of the functional group transforms which add heavy atoms, the isomers of pyridine were considered as substitutes to phenyl (entries 71–73).[31] The pyridine isosteres distributions were all close to a mean of zero with SD similar to the global distribution, with no significance found by matched pair analysis. Pyridines are known to undergo N-oxidation by both flavin-containing mono-oxygenase (FMO) and Cyp450 enzymes,[32] so like the SCH_3 groups discussed above, we may expect a detrimental effect, perhaps more so for the 3 and 4-pyridines where is nitrogen was unhindered (i.e., 2-pyridine would be different). On the opposite hand a reduction in the overall lipophilicity of the compounds may be expected to result in an increase in stability, such that the two effects may have off-set each other. Given these data it is hard to predict more firmly a benefit from these isosteres, but the 2 and 3 isomers did meet the criteria for inclusion in the 'Unlikely to change HLM stability' set.

For each of these transforms we know the change in volume, electronic donation and withdrawing effects from the literature.[33,34] To our disappointment, only rough trends rather than meaningful correlations with either of these individual properties were found. For example, increased volume (A^2) of the 2-position groups tended to lead to a decrease in stability. Similarly no strong correlation was found for Δ Log $D_{7.4}$ for groups at the 2 and 3 positions. However a rough trend exists in the dataset for all the 4-position transforms examined (Chart 3), with some outliers. The spread of data was evident, with groups like 4-$CH(CH_3)_2$, 4-$C(CH_3)_3$ and 4-phenyl having a similar change in lipophilicity but different mean changes in stability. Various models and work have been conducted and published to build QSAR models for individual isoforms,[35–37] and single point mutations for 3A4 for example,[38] and other P450 enzymes.[39–44]

An attempt has been made to provide a pragmatic way of combining the above sets of information using a consensus score (column C—Table 1) to yield improved sets of transforms for use in 'lean design'. The functional transform scores three stars if it had a clear positive effect on increasing HLM stability and appears in both the Topliss set, and Hajduk and Sauer's sets of 24 compounds; two stars for appearing in two of the sets; and one star if the group had only a positive effect on stability. Using the consensus scoring we found six transforms that score 3 stars (4-SO_2CH_3, 4-CF_3, 4-F, 4-Cl, 4-Br, 4-SO_2NH_2) and a further 22 that scored two out of three. These cover both EW and EDG, a range of lipophilicity and volume changes and provide a smaller set of transforms for SAR and DMPK optimisation which we have named the *Consensus* set. At this point we examined the data from substructure searching of the MDDR3D

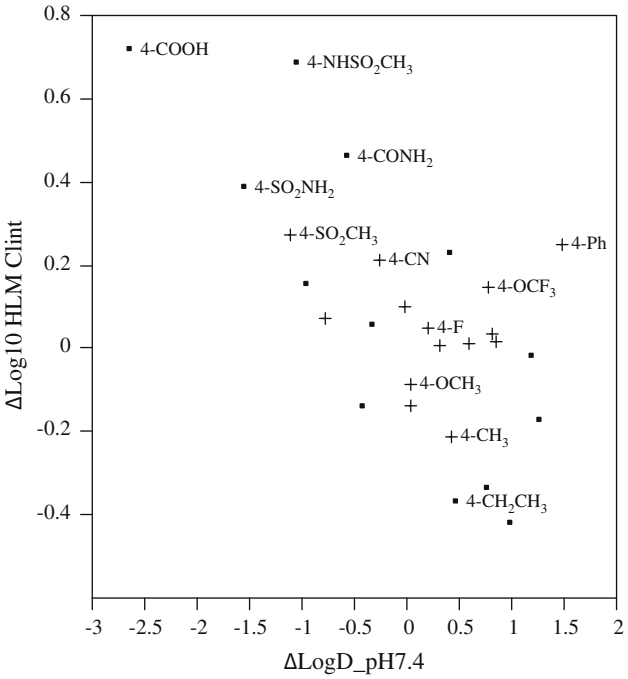


Chart 3. Measured mean ΔLog_{10} HLM Clint stability for each group at the 4-position versus measured Δ Log $D_{7.4}$ + marker $n \geqslant 30$ matched pairs.

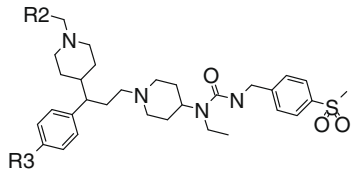
databases to qualitatively corroborate our findings. For example, and only semi-quantitatively, we can say the 4-chlorophenyl occurs frequently in known drugs (45 known drugs out of 390 phenyl bearing drugs), whereas 4-cyano occurs less frequently (5 drugs); but both tend to increase stability. 16 of the 73 groups were found in 6 or more known drugs in the database, and account for 228 of the 390 drugs found. Of these groups, 10 were 'captured' in the consensus set, and 5 in the 'Unlikely to change HLM stability' set, thus in a simple qualitative sense these findings support our overall analysis. However the *consensus* set has only 2 transforms to represent 2-substitution (2-Cl entry 21, and 2-OCH_3 entry 40). This was not surprising given the scoring mechanism was weighted to an increase in HLM stability. We noted above 2-CH_3 was unusual compared to 3 and 4-CH_3, and given the paucity of 2-substituents we propose to the medicinal chemistry community that it is included in the consensus set (Table 4). The exceptions from the known drug searching were 3,4-dimethoxyphenyl (12 known drugs), which has a mean of 0.09 but matched pair analysis $p > 0.05$, and 2,6-dichlorophenyl (9 known drugs), which appeared reliably detrimental but low matched pair numbers, so was not included in any set (mean −0.28, SD 0.16). All other attempts to draw any firmer relationship between these substructure searches of known drugs and our statistical analysis were unsuccessful.

At the time of manuscript preparation a publication by Gleeson, Bravi, Modi and Lowe examined small group matched pairs for P450 inhibition, hERG, solubility and permeability.[45] Their overall conclusion was 'there are no perfect substituents, that is, groups that lead to a significant benefit across all ADMET parameters', and scientists should give 'priority to different scaffolds'. However within their tables of data several groups clearly emerged as beneficial a good proportion of the time [e.g., $C(O)NH_2$, SO_2CH_3, $C(O)CH_3$, $N(CH_3)$, F, CN, OCH_3, OH, CH_2OH].

3. Drug hunting examples

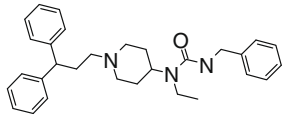
Within the above dataset were a group of fluorine compounds from a CCR5 antagonism project, which illustrate some dramatic

Table 2
Further matched pairs of CCR5 antagonists



Compound	R2 group	R3 group	HLM Clint (μl/min/kg)
3	3-Pyridyl	H	>150
4	3-Pyridyl	F	5.0
5	2-Quinolyl	H	>150
6	2-Quinolyl	F	3.0

Table 3
HLM in vitro stability of CCR5 antagonists



Compound	R1 group(s)	HLM Clint (μl/min/kg)
7	4-H	<4.0 ($n = 2$)
8	4-F	105 ($n = 1$)
9	3-F	86 ($n = 1$)
10	2-F	74 ($n = 1$)
11	2,5-F_2	<4.0 ($n = 1$)
12	3,5-F_2	<3.0 ($n = 1$)
13	2,6-F_2	6.0 ($n = 1$)

examples of changes in HLM stability.[46–48] The matched pair of compounds **3**, **4** and **5**, **6** (Table 2) clearly show the 4-fluoro transform yielding a large increase in stability. This type of example was included in the $F(0.5)$ calculations even though values are out-of-range, because clearly the increase in stability was greater than half a Log unit with confidence (Table 2). Note the 4-SO_2CH_3 group on the right hand phenyl ring.

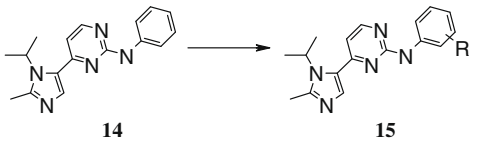
However chemically similar compound **7**, bearing three unsubstituted phenyl rings had a low HLM Clint (Table 3). In these matched pairs addition of a single fluorine at either the 4, 2 or 3 positions (**8**, **9**, **10**) decreased stability, but further substitution (**11**, **12**, **13**) by a second fluorine increased stability back to the original level of compound **7**. This series of compounds raises an interesting problem when interpreting this type of global matched pair data. Were these changes in oxidative metabolism due to blocking or activation of a site, or due to changing the molecular recognition of the compound(s) to certain P450s? Examining compound pairs of **3**, **4** and **5**, **6** it would be easy to conclude that the 4-position of the phenyl was the site of metabolism, and this was blocked. Metabolite identification may even corroborate this, however the compound pairs **7**, **8**; **7**, **9** and **7**, **10** suggest either another site of metabolism or that there was a whole compound recognition event as the structure of **7** was devoid of any substitution. The addition of a fluorine on the right hand ring reduces the electron density and could have increased affinity for a single specific P450, such that an oxidation event could have occurred anywhere on the compound. Indeed compounds **3** and **5** bearing a 4-SO_2CH_3, a more powerful EWG but large and polar, could have increased this recognition. However addition of two fluorines (**11**, **12**, **13**) which again decreased electron donation, stopped the high turnover. This appears inconsistent with this argument, but again had the site of metabolism moved to the left hand phenyl ring? Whatever the case, this illustrates that this type of statistical analysis was clouded by the presence of multiple P450 enzymes and/or binding modes; and while it is possible to use metabolite identification to find the sites, and attempts to block them, it is no guarantee of success.

Compound **14**, a cyclin-dependant kinase 2 (CDK2) inhibitor bearing a phenyl side chain was found to have a moderate/high HLM Clint of 63.5 μl/min/kg. A fair portion of the transforms represented in the consensus set were synthesised in the project (Scheme 3) and Chart 4 illustrates the changes in stability observed. In line with the results above 4-morpholine and 4-SO_2CH_3 gave a clear improvement, and typical of the bulk of results observed in this analysis, most of the transforms did not change in vitro stability by much.[49]

The enzyme, 11-beta hydroxysteroid dehydrogenase type II (11-β-HSD), catalyses the deactivation of glucocorticoids to 11-dehydro metabolites, and compound **16** was a simple ketone inhibitor, with a moderate HLM Clint of 23.0 μl/min/kg.[50] Within the analysis 12 compounds had HLM Clint values, but parallel synthesis was used to make several hundred within the project (Scheme 4—**17**), thereby serving as a test group for the proposed design sets in Table 4. Testing of 48 compounds was completed, and the analysis shown in Charts 5–7 (y-axis Δ Log_{10} HLM Clint). The *consensus* set (Chart 5) yielded 13 out of 25 transforms that

Table 4
Recommended groups of transforms for use in lean design

Design set	Groups	σ para	ΔC Log P (Π)	ΔVol/A^2
Consensus	2-CH_3; 3,4-$(CH_3)_2$; 4-cyclopropyl; 3-CF_3; 4-CF_3; 4-CN; 4-C(O)NH_2; 4-F; 2-Cl; 3-Cl; 4-Cl; 3-Br; 4-Br; 2,4-$(Cl)_2$; 3-CF_3-4-Cl; 2-OCH_3; 4-OCH_3; 2,3-$(OCH_3)_2$; 4-OCF_3; 4-NH_2; 3-$N(CH_3)_2$; 3-NO_2; 4-NO_2; 3-SO_2CH_3; 4-SO_2CH_3; 3-SO_2NH_2; 4-SO_2NH_2; 4-Ph; 4-morpholine	−0.16 to 0.71	−1.59 to 1.82	14.4 to 79.3
Unlikely to change HLM stability	2-CH_3; 3,4-$(CH_3)_2$; 3-CH_3-4-CH_3; 2-CN; 3-CN; 4-C(O)CH_3; 2-F; 3-F; 4-F; 3-Cl; 4-Cl; 3-Br; 3,4-$(Cl)_2$; 2-Cl-4-F; 2,3-$(F)_2$; 3,4-$(F)_2$; 3,5-$(F)_2$; 2,6-$(F)_2$; 2-OCH_3; 3-OCH_3; 4-OCH_3; 2-OCF_3; 3-OCF_3; 2-NH_2; 3-SO_2CH_3; 2-pyridine (isostere); 3-pyridine (isostere)	−0.16 to 0.60	−1.63 to 1.42	0 to 60.4
Increase HLM stability	3,4-$(CH_3)_2$; 4-cyclopropyl; 4-CF_3; 4-CN; 3-C(O)NH_2; 4-C(O)NH_2; 4-CO_2H; 4-F; 4-Cl; 4-Br; 3,5-$(Cl)_2$; 3-CF_3-4-Cl; 2-OCH_3; 4-OCH_3; 2,3-$(OCH_3)_2$; 2,4-$(OCH_3)_2$; 3,4-$(OCH_3)_2$; 4-OCF_3; 4-NH_2; 3-$NHSO_2CH_3$; 4-$NHSO_2CH_3$; 3-NO_2; 4-NO_2; 3-SO_2CH_3; 4-SO_2CH_3; 3-SO_2NH_2; 4-SO_2NH_2; 4-Ph; 4-morpholine	−0.16 to 0.71	−1.82 to 1.96	14.4 to 79.3



Scheme 3. CDK2 inhibitor project examples.

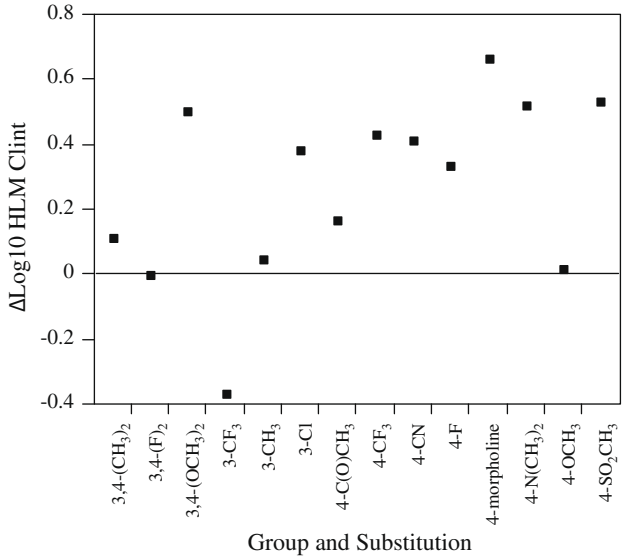
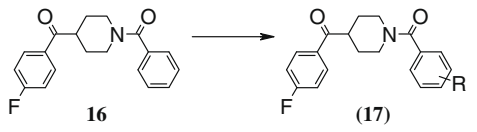


Chart 4. Mean Δ Log_{10} HLM Clint for transform used on CDK2 compounds (**15**). Most small groups gave little change (near to zero line marked) in HLM stability however 4-morpholine and 4-SO_2CH_3 gave an improvement >0.5 Log units.



Scheme 4. 11-β-HSD inhibitor project examples.

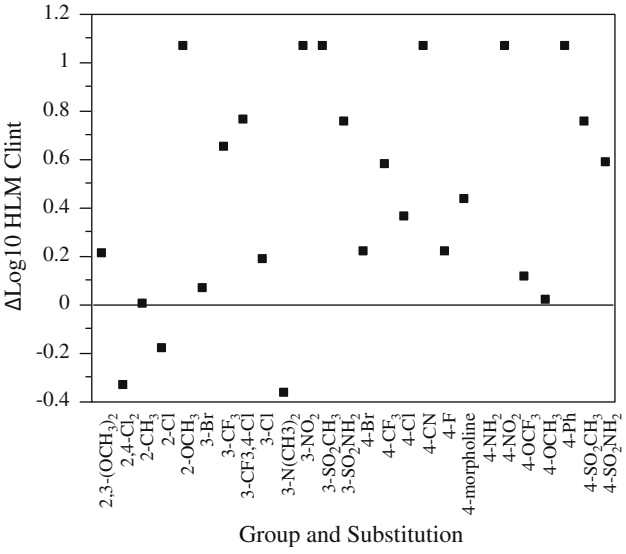


Chart 5. Mean Δ Log_{10} HLM Clint for transforms within the *consensus* design set for 11-β-HSD compounds (**17**). 13 of 25 transforms increase stability with 6 changes reaching the limit of the assay (<2.0 μl/min/kg. thus points at the top of plot). [3,4-$(CH_3)_2$; 4-cyclopropyl; 4-C(O)NH_2; 3-N$(CH_3)_2$ were not synthesised].

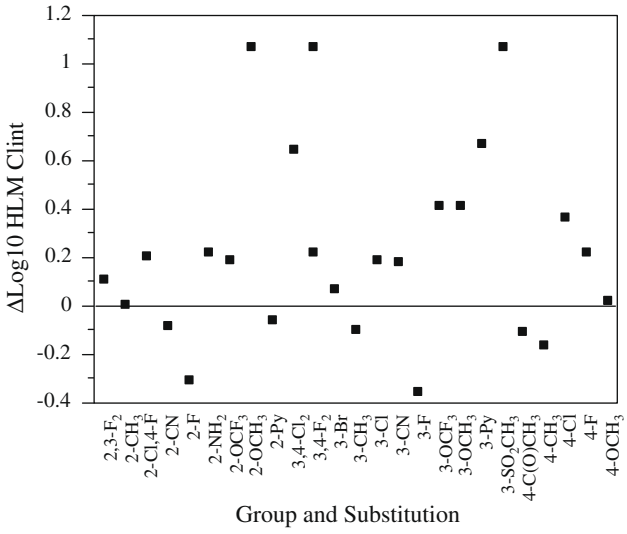


Chart 6. Mean Δ Log_{10} HLM Clint for transforms within the *unlikely to change HLM stability* design set for 11-β-HSD compounds (**17**). 5 of 26 transforms increase stability with 3 changes reaching the limit of the assay (<2.0 μl/min/kg). [3,4-$(CH_3)_2$; 4-cyclopropyl; 2,6-$(F)_2$ were not synthesised]

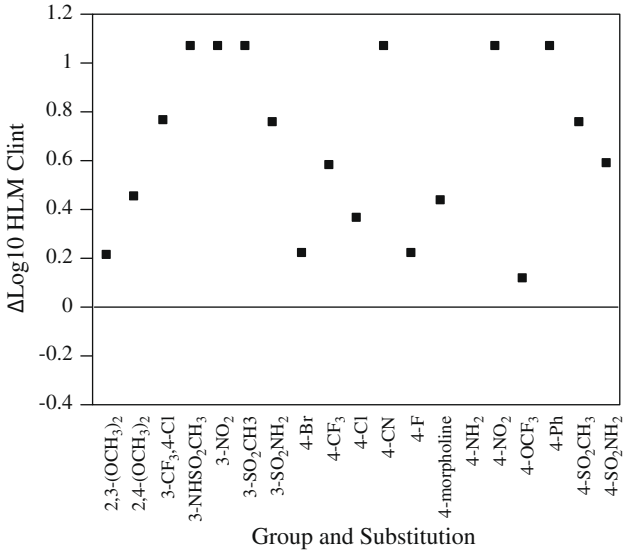


Chart 7. Mean Δ Log_{10} HLM Clint for transforms within the *increase HLM stability* design set for 11-β-HSD compounds (**17**). 12 of 19 transforms increase stability with 6 changes reaching the limit of the assay (<2.0 μl/min/kg). [3,4-$(CH_3)_2$; 4-cyclopropyl; 3-C(O)NH_2; 4-C(O)NH_2; 4-CO_2H; 3,5-$(Cl)_2$; 3,4-$(OCH_3)_2$; 3-NH_2; 4-NH_2; 4-$NHSO_2CH_3$ were not synthesised].

increased HLM stability by more than half a Log unit to the limit of the assay. The transforms 4-F, 4-Ph, 4-NO_2, 4-Cl and 4-Br all maintained affinity for the enzyme with concomitant increase in stability. A lower portion of the transforms within the *unlikely to change HLM stability* set increased stability as expected, although 3 transforms had stability past the limit of the assay (Chart 6). Of these 2-OCH_3 and 3-SO_2CH_3 are within the *consensus* set so potentially the transform 3,4-$(F)_2$ could have been overlooked using only the *consensus design set*. Lastly the *increase HLM stability* set yielded strong results with 12 out of 19 increasing stability by greater than half a Log unit (Chart 7).

4. Conclusions

The analysis of in vitro human liver microsomal turnover, combined with substituents on known drugs, has revealed a group of simple functional transforms that can be added to phenyl rings with a high probability of having little effect on oxidative metabolic liability. This information is of highest utility when the medicinal chemist wishes to alter other compound properties without detriment (Table 4). The global dataset also revealed that changes in stability of >1.0 Log unit are possible, but unlikely for most small group additions, and many of these changes are equally likely to have a detrimental effect. In this scenario a medicinal chemist faced with a series of compounds, all with high metabolic turnover, is unlikely to overcome this issue; and if combined with several issues may wish to work in alternative areas. We have used a lower cut-off of 0.5 Log unit to provide a pragmatic way forward and meaningful frequencies of likelihood to increase or decrease stability for use in molecular design, as a threefold increase in stability when scaled in vivo may make all the difference in optimisation of a lead into candidate drug. Given the nature that molecular recognition plays in in vitro metabolism, and the mixture of cytochrome P450 enzymes in microsomal extracts, it must be acknowledged that the effects of substituent properties has been 'diluted out' within the spread of data, hence a rough correlation with lipophilicity was observed. Nonetheless the data re-enforced some of the perceptions of medicinal chemists and those substituent effects already known in the literature. For example these data show that iso-propyl and sulfide groups have high turnover, and most importantly highlight again those groups (4-F, 4-CN, 4-SO_2CH_3) which are known to increase stability. Surprisingly, these groups are not highly represented in known drugs, but the simple groups such as methyl, chloro, fluoro and methoxy, tended to dominate and were captured in the design sets. Added to this the fact that 277 known drugs have a pendant phenyl suggests that the overall size, shape and properties of the compound will dominate its stability in the end (the same conclusion was reached by Gleeson et al. for other DMPK properties).[42] Whilst imperfect, by using a consensus scoring we propose 29 transforms that are more likely to increase HLM stability (summarised in Table 4), increase affinity to target receptor and cover diverse property space. We propose two further sets of transforms where maintaining HLM stability is desirable and where increase in stability alone is desired. The wider dataset in Table 1 presented here should help the medicinal chemist in selecting transforms to aid pragmatic lean design, especially where chemical synthesis is difficult and only a limited number of compounds can be made. Overall given these data, previous work from the literature, known drugs and 'Topliss thinking' it should now be possible to synthesise fewer compounds to yield improvements, or, ever more importantly, rapid stop decisions.

5. Computational procedures and matched pair analysis

Matched pairs for a given transform were found from the global in vitro HLM dataset by using in-house proprietary software ('ThricePairs'), to yield a total of 6051 matched pairs. Out-of-range values were removed for statistical evaluation yielding 5321 matched pairs. Compound diversity was examined by considering the mean Tanimoto distance across each group, which was determined by in-house program 'alfi' and the Tanimoto distances calculated using the program 'snailflush'.[51–53] Only four matched pair groups were found to have a mean Tanimoto below 0.6 (where 0.0 maybe an exact match and 1.0 highly diverse) and all data have been included in Table 1 as there may still be useful information to interpret from them. The searches conducted here for specific groups at specific positions on a phenyl ring have resulted in low numbers of matched pairs for some transforms (35 of the 73 transforms have <30 matched pairs). However the majority of these have high chemical diversity as indicated from the mean Tanimoto distance, so have been included in Table 1 and in all further analysis (on plots these are marked crosses). Matched pair analysis was performed using SAS JMP version 6.0 and *p*-values reported in Table 1. These corresponded with the mean Δ Log_{10} HLM Clint greater than 2 × SE for the distribution. There are several methods available for comparing distribution data; Dunnet[54,55] method involves selecting a distribution as a point of reference and comparing the overlap of the rest to it; Tukey and Kramer[56] method performs all of the combinations as comparisons and was used to compare all the fluorine substitutions. All of the methods produced the same results as matched pair analysis within each transform. As described above the total dataset of 6051 matched pairs was used to find frequency greater than or equal to 0.5 Δ Log_{10} HLM Clint increase in stability [*F*(0.50)] and less than −0.5 Δ Log_{10} HLM Clint (i.e., decrease in stability) [*F*(−0.50)]. The subtraction of Log terms in this fashion (Log_{10} (Ph) − Log_{10} (Sub-Ph)) was performed such that an increase in stability was a positive value to aid clarity for the reader. After examining several 'cut-offs', and considering the proposals and probabilities published by Lewis et al. using a >2-fold level, we proposed that 0.5 Log_{10} unit of change in HLM level was reasonable (i.e., 3.2-fold), and in many cases had sufficient populations to be studied by 2 × 2 contingency testing to yield low *p*-values.[57] It is important to point out this was a mathematical testing of population relative to the 4-fluoro transform distribution, and made the assumption that this set was robust and had minimised errors. In other words how likely are tails of the distribution real for the given transform compared to 4-fluoro? In Table 1 we have published the frequency data and 2 × 2 testing *p*-values, noting the assumptions above. In the same manner that the transform 4-F-Ph to 4-CN-Ph was studied (matched pair C, Scheme 2) to examine the comparability of transform distributions the following were also tested: [4-F-Ph to 4-CH_3SO_2-Ph, n = 28, mean 0.28, SE 0.09, p = 0.002]; [4-F-Ph to 4-CH_3-Ph, n = 85, mean −0.30, SE 0.06, p <0.0001]; [4-F-Ph to 3-pyridine, n = 54, mean −0.07, SE 0.06, p >0.05], indicating that the data here were comparable between transforms from phenyl.

Acknowledgements

The author would like to acknowledge the valuable discussions with Drs. E. Griffen, J. Cumming, M. Waring, A. Leach, P. Kenny, P. Ballard and R. Butlin, and all of the scientists across AstraZeneca that tested the compounds used in this analysis.

References and notes

1. Topliss, J. G. *J. Med. Chem.* **1972**, *15*, 1006.
2. Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Disc.* **2007**, *6*, 881.
3. Barnhart, T. *J. Compet. Lean Think Future State* **2008**, *spring*, 1.
4. Russell, K. *Pharma Focus Asia* **2008**, *7*, 48.
5. Andersson, S.; Armstrong, A.; Bjore, A.; Bowker, S.; Chapman, S.; Davies, R.; Donald, C.; Egner, B.; Elebring, T.; Holmqvist, S.; Inghardt, T.; Johannesson, P.; Johansson, M.; Johnstone, C.; Kemmitt, P.; Kihlberg, J.; Korsgren, P.; Lemurell, M.; Moore, J.; Pettersson, J. A.; Pointon, H.; Ponten, F.; Schofield, P.; Selmi, N.; Whittamore, P. *Drug Discovery Today* **2009**, *14*, 598.
6. Hammond, C.; O'Donnell, C. J. *Drug Discovery World* **2008**, *spring*, 9.
7. Dossetter, A.G.; Cook, C.; Morris. J., unpublished results. Data for compound 1. LCMS R_t = 1.35 min $[M+H]^+$ found 366.37; 368.34 required $C_{19}H_{21}Cl_2NO_2$ 366.29; ^{1}H NMR (400.13 MHz, $CDCl_3$, 303K) δ 0.85–1.00 (m, 2H); 1.11 ('q', 2H); 1.39 ('hp', 2H); 2.31 (q, 1H); 3.26 (dd, 2H); 3.30 (q, 1H); 4.50 (s, 2H); 6.27 (d, 1H); 6.64 (d, 1H); 6.75–6.85 (m, 5H).
8. Scientists at Berlex laboratories also discovered a similar chemical series, and similar SAR: Wei, R. G.; Arnaiz, D. O.; Chou, Y.-L.; Davey, D.; Dunning, L.; Lee, W.; Lu, S.-F.; Onuffer, J.; Ye, B.; Phillips, G. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 231.
9. Hajduk, P. J.; Sauer, D. R. *J. Med. Chem.* **2008**, *51*, 553.
10. Lewis, M. L.; Cucurull-Sanchez, L. *J. Comput. Aided Mol. Des.* **2009**, *23*, 97.
11. A typical procedure for the HLM Clint assay was as follows: The compounds to be incubated are added from a concentrated stock (e.g., 0.2–1.0 μM) in an appropriate solvent (e.g., DMSO, 1% v/v) to a suspension of microsomal protein

(final concentration 0.5–2.0 mg ml^{-1}) in a suitable vial. After a 2 min pre-incubation at 37 °C, the cofactor NADPH is added (final concentration of 1.0 mM) and the reactions allowed to proceed. The various additions are made using automated processes (Quadra 96 multi-pipetting workstation). At appropriate time points (e.g., 0, 5, 10, 20 and 30 min), an aliquot (e.g., 100 μl) is taken from the incubation and added to 2–3 vols of acetonitrile to terminate the reactions and denature the microsomal enzymes. Control incubations may also be conducted from which NADPH or compound have been omitted. Once the incubations have been quenched, the samples are shaken for 5 min and then centrifuged for 15 min at 3000 rpm and 4 °C. The supernatants are taken and analysed by HPLC using the method developed for the particular project or the probe substrates of interest. The following HPLC conditions are typical: solvents: A: methanol and B: 0.025% ammonium acetate in water (w/v), Column: Waters Symmetry C_8 75 × 4.6 mm, 3.5 μm, flow rate 1.5 ml min^{-1} gradient: 5–95% A over 7 min, held at 95% for 1.5 min.

12. For data analysis methods used at AstraZeneca (PLC) and a similar process examining physiochemical processes see: 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.
13. The MDL Drug Data Report, available from Symyx Technologies. http://www.symyx.com/products/databases/bioactivity/mddr/index.jsp (accessed and searched Dec 1st 2008).
14. Sheridan, R. P. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 103.
15. For a recent review of the role of fluorine in the metabolism of Drug molecules see: Park, K. B.; Kitteringham, N. R.; O'Neill, P. M. *Annu. Rev. Pharmacol. Toxicol.* **2001**, *41*, 443.
16. For a discussion of the role of fluorine in medicinal chemistry see: Hagmann, W. K. *J. Med. Chem.* **2008**, *51*, 4359.
17. Coon, M. J.; Vaz, A. D. N.; McGinnity, D. F.; Peng, H.-M. *Drug Metab. Dispos.* **1998**, *26*, 1190.
18. Guengerich, F. P.; MacDonald, T. L. *FASEB* **1990**, *4*, 2453.
19. Wade, R. C. *J. Comput. Aided. Mol. Des.* **1990**, *4*, 199.
20. Guengerich, F. P.; Yun, C.-H.; MacDonald, T. L. *J. Biol. Chem.* **1996**, *271*, 27321.
21. Smith, D. A.; Ackland, M. J.; Jones, B. C. *Drug Discovery Today* **1997**, *2*, 406.
22. Smith, D. A.; Ackland, M. J.; Jones, B. C. *Drug Discovery Today* **1997**, *2*, 479.
23. Smith, D. A.; Jones, B. C.; Walker, D. K. *Med. Res. Rev.* **1996**, *16*, 243.
24. Koerts, J.; Soffers, A. E. M. F.; Vervoort, J.; De Jager, A.; Rietjens, I. M. C. *Chem. Res. Toxicol.* **1998**, *11*, 503.
25. Dear, G. J.; Ismail, I. M.; Mutch, P. J.; Plumb, R. S.; Davies, L. H. *Xenobiotica* **2000**, *30*, 407.
26. Cnubben, N. H. P.; Peelen, S.; Borst, J. W.; Vervoort, J.; Veeger, C. *Chem. Res. Toxicol.* **1994**, *7*, 590.
27. Smith, D. A.; Jones, B. C. *Biochem. Pharmacol.* **1992**, *44*, 2089.
28. Applequist, D. E.; Klug, J. H. *J. Org. Chem.* **1978**, *43*, 1729.
29. Ferguson, K. C.; Whittle, E. *Trans. Faraday Soc.* **1971**, *67*, 2618.
30. Chauret, N.; Guay, D.; Li, C.; Day, S.; Silva, J.; Blouin, M.; Ducharme, Y.; Yergey, J. A.; Nicoll-Griffith, D. A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2149.
31. Additional heterocycle isosteres have been studied and these results will be published in due course.
32. For an example of N-oxidation of pyridine with a drug compound see: Bu, H.-Z. *Drug Metab. Dispos.* **2006**, *34*, 1798.
33. Hansch, C.; Leo, A.; Hoekmann, D. *J. Am. Chem. Soc.* **1995**, *2*.
34. Kubinyi, H. *Quant. Struct.-Act. Relat.* **1988**, *7*, 121.
35. Lewis, D. F. V.; Dickins, M. *Toxicology* **2003**, *170*, 45.
36. Lewis, D. F. V.; Dickins, M. *Drug Discovery Today* **2002**, *7*, 918.
37. Lewis, D. F. V.; Lake, B. G. *Toxicology* **1998**, *125*, 31.
38. For QSAR and enzyme research on P450 3A4 specific: Szklarz, G. D.; Halpert, J. R. *J. Comput. Aided. Mol. Des.* **1997**, *11*, 265.
39. Donamski, T. L.; Liu, J.; Harlow, G. R.; Halpert, J. R. *Arch. Biochem. Biophys.* **1998**, *350*, 223.
40. Ekins, S.; Bravi, G.; Wikel, J. H.; Wrighton, S. A. *J. Pharmacol. Exp. Therap.* **1999**, *291*, 424.
41. Harlow, G. R.; Halpert, J. R. *J. Biol. Chem.* **1997**, *272*, 5396.
42. He, Y. A.; He, Y. Q.; Szklarz, G. D.; Halpert, J. R. *Biochemistry* **1997**, *36*, 8831.
43. Afzelius, L.; Zamora, I.; Ridderstrom, M.; Andersson, T. B.; Karlen, A.; Masimirembwa, C. M. *Mol. Pharm.* **2001**, *59*, 909.
44. Lewis, D. F. V. *Xenobiotica* **2002**, *32*, 305.
45. Gleeson, P.; Bravi, G.; Modi, S.; Lowe, D. *Bioorg. Med. Chem. Lett.* **2009**, *17*, 5906.
46. Burrows, J. N.; Cumming, J. G.; Fillery, S. M.; Hamlin, G. A.; Hudson, J. A.; Jackson, R. J.; McLaughlin, S.; Shaw, J. S. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 25.
47. Cumming, J. G.; Cooper, A. E.; Grime, K.; Logan, C. J.; McLaughlin, S.; Oldfield, J.; Shaw, J. S.; Tucker, H.; Winter, J.; Whittaker, D. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5012.
48. Cumming, J. G.; Brown, S.; Cooper, A. E.; Faull, A.; Flynn, A.; Grime, K.; Oldfield, J.; Shaw, J. S.; Shepard, E.; Tucker, H.; Whittaker, D. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 3533.
49. Anderson, M.; Andrews, D. A.; Barker, A. J.; Brassington, C. A.; Breed, J.; Byth, K. F.; Culshaw, J. D.; Finlay, M. R. V.; Fisher, E.; McMiken, H. H. J.; Green, C. P.; Heaton, D. W.; Nash, I. A.; Newcombe, N. J.; Oakes, S. E.; Paupit, R. A.; Roberts, A.; Stanway, J. J.; Thomas, A. P.; Tucker, J. A.; Walker, M.; Weir, H. M. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5487.
50. Barton, P. J.; Jewesbury, P. J.; Pease, J. E. WO 2004/033427.
51. Blomberg, N.; Cosgrove, D. A.; Kenny, P. W.; Kolmodin, K. *J. Comput. Aided Des.* **2009**, *23*, 513.
52. For recent comparison of similarity methods and references within see: Haranczyk, M.; Holliday, J. *J. Chem. Inf. Model.* **2008**, *48*, 498.
53. Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. *J. Med. Chem.* **1996**, *9*, 3049.
54. Dunnett, C. W.; Crisafior, J. *J. Pharm. Pharmacol.* **1955**, *7*, 314.
55. Dunnett, C. W. *J. Am. Stat. Ass.* **1955**, *50*, 1096.
56. Tukey, 1953, unpublished proposal: Kramer, C.Y. *Biometics*, **1956**, *12*, 307.
57. In other words >1 Log_{10} unit was not populated enough, and more likely to be error prone, and as low as twofold was too close to experimental error of the assay.