The influence of drug-like concepts on decision-making in medicinal chemistry

Paul D. Leeson and Brian Springthorpe

Abstract | The application of guidelines linked to the concept of drug-likeness, such as the 'rule of five', has gained wide acceptance as an approach to reduce attrition in drug discovery and development. However, despite this acceptance, analysis of recent trends reveals that the physical properties of molecules that are currently being synthesized in leading drug discovery companies differ significantly from those of recently discovered oral drugs and compounds in clinical development. The consequences of the marked increase in lipophilicity — the most important drug-like physical property — include a greater likelihood of lack of selectivity and attrition in drug development. Tackling the threat of compound-related toxicological attrition needs to move to the mainstream of medicinal chemistry decision-making.

LogP
Log of the octanol–water partition coefficient, which is a measure of a drug's lipophilicity. Defined as the ratio of un-ionized drug distributed between the octanol and water phases at equilibrium. Higher values imply greater lipophilicity.

Molecular mass
The molecular mass of a substance, frequently called molecular weight, is the mass of one molecule of that substance, and its units are the unified atomic mass unit (u) or Dalton (Da), which equals 1/12 the mass of one atom of carbon-12.

Polar surface area
(PSA). This is defined as the surface sum over all polar atoms, (usually oxygen and nitrogen), also including attached hydrogens.

AstraZeneca R&D Charnwood, Bakewell Road, Loughborough LE15 5RH, UK.
Correspondence to P.D.L.
e-mail:
paul.leeson@astrazeneca.com
doi:10.1038/nrd2445

Ten years have passed since the publication of the 'rule of five' physical property guidelines for drug permeability[1,2]. The rule of five, which was derived from a database of clinical candidates reaching Phase II trials or further, states that poor absorption or permeability are more likely when cLogP (the calculated 1-octanol–water partition coefficient) is >5; molecular mass is >500 daltons (Da); the number of hydrogen-bond donors (OH plus NH count) is >5; and the number of hydrogen-bond acceptors (O plus N atoms) is >10 (REFS 1,2). Although the rule-of-five properties are interrelated (cLogP actually being a composite property dependent on molecular size, polarity and hydrogen bonding[3]), its conceptual simplicity and ease of calculation has made it the leading measure of drug-likeness, with the original article having more than 1,500 literature citations.

The physicochemical profiles of oral drugs[4–6] are consistent with the rule of five, and recent studies on drug absorption have also highlighted the importance of polar surface area (PSA), which is closely correlated to O plus N atom count, and LogD (1-octanol–water coefficient at various pH values)[7,8]. Overall oral bioavailability depends not just on absorption but also on dissolution, gut transit time and first-pass metabolism, so it is perhaps surprising that rat bioavailability can be categorized by simple physical properties such as PSA and rotatable bonds[9,10], and number of rule-of-five violations, PSA and ionization state[11]. However, these models do not readily explain human bioavailability data[12]. One further aspect that has received less attention is that bulk molecular properties are relevant to drug safety as well as pharmacokinetics and metabolism: marketed oral drugs, which are rule-of-five compliant[4–6], have by definition successfully passed rigorous toxicological and clinical safety hurdles.

Analyses of drug-likeness inevitably depend on the medicinal chemistry innovation and prevailing strategies of the past. However, it is clear that compounds produced by more recent medicinal chemistry efforts do not occupy the same chemical space as historical drugs. From three studies of oral drugs[4–6], the mean cLogP values were 2.3, 2.5 and 2.5 and mean molecular masses were 333, 337 and 344 Da. By contrast, 1,117 GlaxoSmithKline compounds[9] and 553 Abbott compounds[11] that had advanced to pharmacokinetic studies had, respectively, mean cLogP values of 4.3 and 3.9, and molecular mass values of 480 and 434 Da. Additionally, a group of 1,680 optimized compounds from the recent medicinal chemistry literature had a mean cLogP of 4.0 and molecular mass of 435 Da[13]; and in another literature study, more than 50% of compounds with high potency had cLogP values >4.25 and molecular mass values >425 Da[14]. Drugs and research compounds also differ in physicochemical properties between target protein classes[13,15,16] and therapy areas[17]. Molecular mass and other properties of oral drugs[6,17], as well as of literature compounds[16], are also increasing with time.

These changes are potentially a concern because physical property inflation may prove detrimental to the health of drug development pipelines. Several studies[4,14,16,18] of compounds in development concur that mean molecular mass declines as compounds progress through Phase I, II and III; moreover, the more

lipophilic compounds tend to be discontinued at each phase[4]. These observations are likely to be the result of developmental selection pressures, combined with the increase in physical properties of compounds entering drug development.

With a mean time of ~12 years from drug discovery to launch, any application of drug-like concepts since the late 1990s will have only had a marginal influence on today's late-stage development pipelines. However, there has been ample opportunity for drug-likeness to have an impact

Box 1 | **Data collection and analysis**

Oral drugs. An oral drugs database containing 2,118 drugs approved worldwide up to June 2007 was assembled from the Food and Drug Administration Orange book[56] and published databases[5,6,57]. For oral drugs launched since 1983 (592 drugs), the year of launch was collected from literature compilations[57,58] and the year of first publication of the drug (usually a patent[6]) was obtained from Scifinder[59] and the Merck Index[60].

Promiscuity data. The Cerep BioPrint database ('BioPrint' is a registered trademark of Cerep SA)[27,28] contains data on 2,133 compounds, mainly existing drugs and reference compounds, which have been tested in an *in vitro* panel of 200 receptor, enzyme and metabolic screens. Promiscuity is defined as the number of hits for each compound, for which >30% inhibition was found at a concentration of 10 μM; compounds meeting this level of activity had IC_{50} values determined. This is a larger data set than the BioPrint data examined previously (1,098 compounds)[29].

Development compounds. Compounds originating with, or associated with, the top 25 pharmaceutical companies[61] and in Phase I or I–II (n = 91), Phase II or II–III (n = 214), and Phase III or preregistration (n = 126) were collected from the Prous Science Integrity database ('Integrity' is a registered trademark of Prous)[57] in July 2007.

Patented compounds. Databases of the most recently published compounds in patent applications originating from four major pharmaceutical companies — AstraZeneca, GlaxoSmithKline, Merck and Co., and Pfizer — were curated from two sources, Prous Science Integrity[57] and GVK Bio[62]. Prous Science Integrity data were collected from January 2001 to June 2007; over this time period the four selected companies produced the largest number of patent applications compared with other organizations. Prous Science Integrity data include compounds from legacy organizations now merged with the selected companies (Merck Frosst, and Merck Sharp and Dohme are abstracted separately and were added to the Merck and Co. data). Compounds with molecular mass <1,000 daltons were selected from the earliest phase, defined as 'biological testing' in Prous Science Integrity, which removes patents on existing drugs and development compounds. The data were further filtered, by selecting 'lead compounds' only — the Prous Science Integrity abstractors selected (mostly) a single compound from the patent application, either a specifically claimed compound, or the example with the best cited biological activity, or a compound representative of the others in the patent. Finally, basic patent applications, filed by the PCT route (Patent Cooperation Treaty for international patent applications, 'WO' numbered patents filed from 2001 onwards, >90% of the total) were selected, providing the year of publication. This resulting database contained 3,516 entries from 3,389 WO patents (855 AstraZeneca, 921 GlaxoSmithKline, 822 Merck and 918 Pfizer entries).

Unlike Prous Science Integrity, the available GVK Bio database is limited to compounds patented in selected target classes only: kinases, GPCRs (G-protein-coupled receptors), proteases, ion channels, transporters, nuclear hormone receptors and phosphatases. GVK Bio abstractors do not select specific compounds; instead a comprehensive coverage (sometimes limited to 100–200 compounds for a patent application) is provided. The database was searched for WO patents from the four selected companies for the period 2003–2007. Legacy and associated organizations appeared to be included separately in GVK Bio. Therefore the search terms used included the names Merck & Co, Merck and Co, Merck Frosst, and Merck Sharp for Merck; Glaxo and SmithKline for GlaxoSmithKline; Pfizer, Warner–Lambert, Upjohn and Pharmacia for Pfizer. AstraZeneca was used as a single search term. Some patent duplications under the different legacy organizations were found and removed. The numbers of compounds per patent from these searches ranged from one to several hundred; to provide a balanced assessment of each patent, the mean and median physical property data per patent were used in addition to the total numbers of compounds. Two GVK Bio databases resulted: GVK compounds containing 117,148 compounds (AstraZeneca = 25,233, GlaxoSmithKline = 28,810, Merck = 31,327 and Pfizer = 31,778), and GVK patents containing 1,903 WO patents (AstraZeneca = 440, GlaxoSmithKline = 497, Merck = 519 and Pfizer = 447).

In each database, GPCR targets were assigned according to the endogenous ligand — that is, amine, peptide or lipid[15,63]. Kinase, protease, ion channel, transporter and nuclear hormone receptor target classes were also assigned to the Prous Science Integrity compounds, based on the abstracted mechanism of action. The GVK Bio assignments of these target classes were used as provided. The numbers of compounds and patents and 1983–2007 oral drugs in each target class examined is summarized in Supplementary information S4 (table). A random selection of chemical structures and patent assignments from the assembled databases were checked for accuracy.

Data analysis. Physical property data were obtained from AstraZeneca's C-Lab tool, incorporating standard packages for LogP calculations (cLogP, ACDLogP), and an in-house algorithm for the distribution coefficient (1-octanol–water LogD at pH 7.4). Other well-established molecular properties[4–6,9,11] examined were polar surface area (PSA), the hydrogen-bond donor and acceptor counts, the numbers of rings and rotational bonds, the Lipinski score (a value of 1 for each violation of the rule of 5 per molecule, with maximum value of 4), and the charge state (acid, base, quaternary base, neutral or zwitterion). Statistical analyses were performed with JMP[64] and visualized with Spotfire[65]. In line with recent practice[13] we cite median rather than mean physical property values for normally distributed continuous data (molecular mass and cLogP), although the use of mean data does not materially change the data or conclusions reached. For full statistical details, including promiscuity analyses and the results of paired Student's t-tests between companies from JMP, see Supplementary information S5 (box).

Bioavailability
This is the fraction of an oral dose that reaches the systemic circulation.

Chemical space
This is the space spanned by all energetically stable stoichiometric combinations of electrons, atomic nuclei and topologies in molecules. Drug-like space may contain 1×10^{20} to 1×10^{200} molecules. All these molecules can never be made — to date 2.7×10^{7} molecules have been reported.

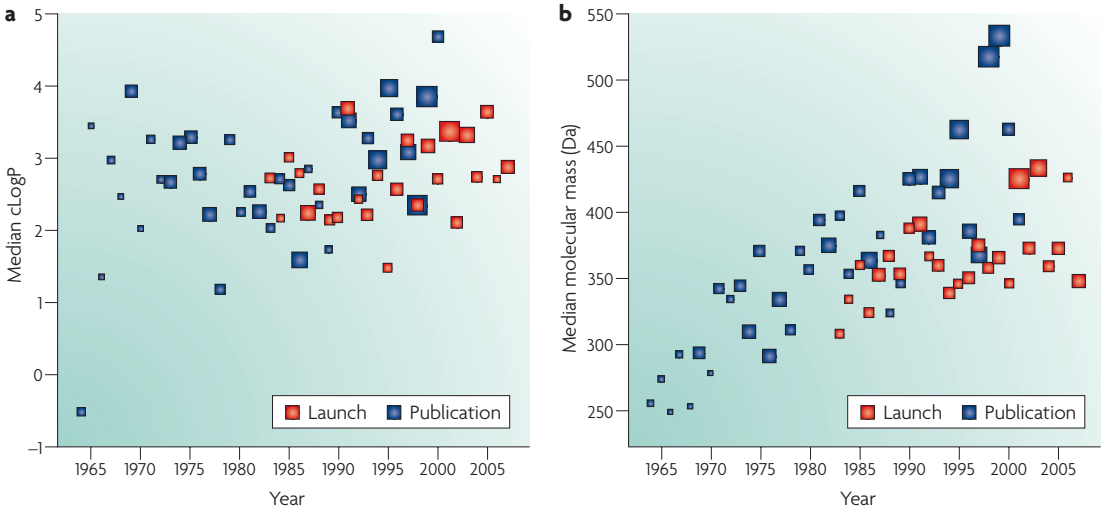

Figure 1 | **Trends in cLogP and molecular mass in launched drugs.** The relationships between median cLogP (**a**) or molecular mass (**b**) and launch or publication dates for 592 oral drugs approved worldwide between 1983 and 2007 are shown. The publication data exclude 29 drugs published pre-1964 or post-2001. Launch and publication dates differ by a median of 10.5 years, consistent with drug development time frames. For each launch year, 10–100% of the drugs occupied a median publication time of 20.6 years. Hence, using publication dates[6] gives a better view of contemporary practice than launch dates. The results of straight line fits are: molecular mass publication: $r^2 = 0.70$, slope = 5.2, $p < 0.0001$; molecular mass launch: $r^2 = 0.27$, slope = 2.1, $p = 0.0076$; cLogP publication: $r^2 = 0.17$, slope = 0.034, $p = 0.010$; cLogP launch: $r^2 = 0.095$, slope = 0.022, $p = 0.13$. The size of the squares represents the mean Lipinski score with range 0 (smallest) to 1 (largest) (see also Supplementary information S5 (box) panels 1,2).

on current decision-making in medicinal chemistry. To examine this, we decided to compare the physicochemical profiles of recently discovered oral drugs and compounds in development with the most contemporary biologically active compounds: those published in the current patent literature (for a description of data collection and analysis, see BOX 1). We chose to look at patent applications for small molecules originating from four large multinational organizations: AstraZeneca, GlaxoSmithKline, Merck and Co., and Pfizer. These companies are the most prolific contributors of new patent applications in this field and have a spread of therapeutic interests covering most areas of current drug discovery, as well as a wealth of know-how in drug discovery and development. We also examine the relationship between physical properties and drug promiscuity, one possible source of toxicological side effects.

Oral drugs

In oral drugs approved since 1983, there are increases with publication date[6] in three of the four rule-of-five properties (molecular mass, O plus N atom count and OH plus NH count), whereas one (lipophilicity, cLogP) is changing less appreciably (molecular mass and cLogP data are shown in FIG. 1). Median molecular mass increased significantly on average by 5.2 Da per publication year from 1964–2001, and by 2.2 Da per launch year from 1983–2007. Given the large increase in molecular mass and small change in cLogP, it is not surprising that other related bulk physical properties — that is, PSA, rings and rotatable bond count — are also increasing significantly with publication time, whereas other measures of lipophilicity (AZLogD7.4) and percentage of PSA[17] are not (Supplementary information S1 (table)).

Factors driving the increases in physical properties in the most recently discovered oral drugs include a greater number of apparently less druggable new targets[19,20], for which larger and more lipophilic molecules appear to be necessary for high affinity binding to active sites — for example, HIV protease inhibitors and antagonists of peptidic G-protein-coupled receptors (GPCRs) — and intensive optimization of specific molecular classes

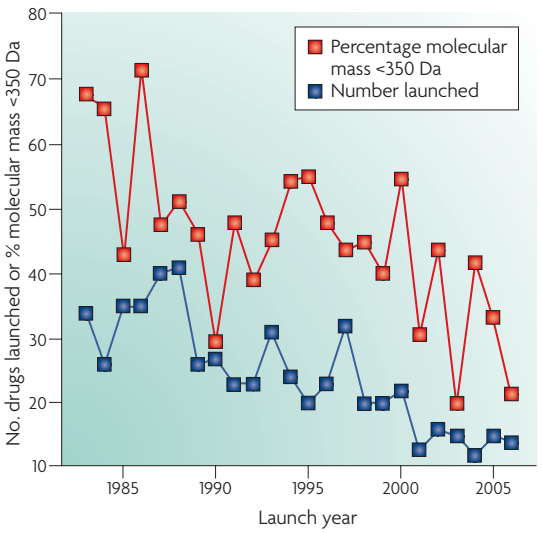

Figure 2 | **Trends in drug approvals and their molecular mass.** The number of oral drugs approved per annum worldwide from 1983 to 2006 and the percentage of these drugs that have molecular mass <350 Da are shown.

with common substructures or pharmacophores[17]. For example, using publication dates from Proudfoot[6], there are time-related increases in molecular mass with year of publication in benzodiazepines, β-adrenoceptor antagonists, histamine H_2 antagonists, dihydropyridine calcium-channel blockers, non-steroidal anti-inflammatory drugs and quinolone antibiotics (Supplementary information S2 (box)).

The reduction in the proportion of launched low-molecular-mass oral drugs over time correlates with the established decline in new drug launches[21] (FIG. 2). Although we do not propose that this provides a complete explanation of the drop in productivity experienced by the pharmaceutical industry, as there are several other possible reasons[21], the possibility that it has been an important contributory factor deserves serious consideration.

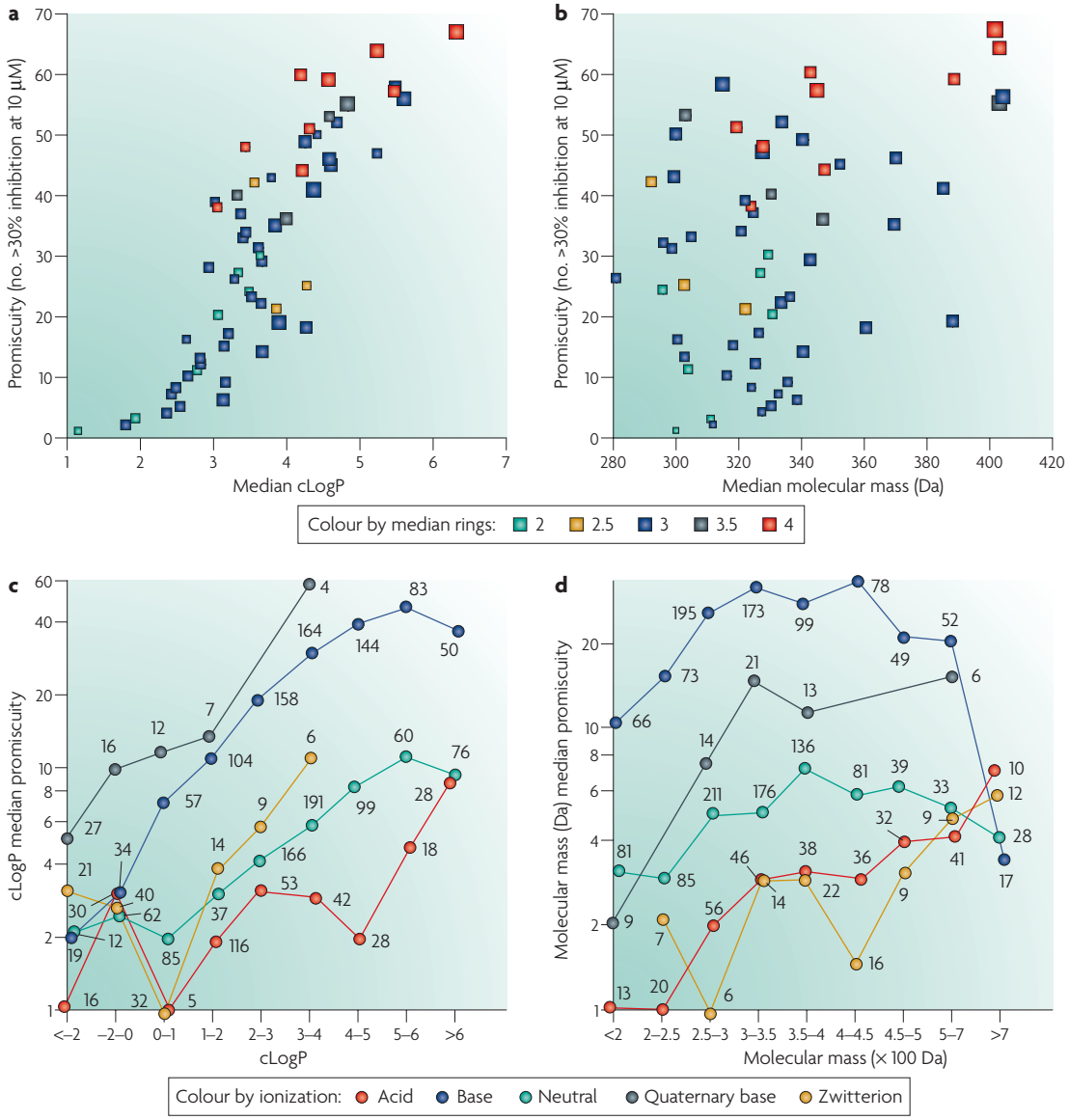

Figure 3 | **Promiscuity analysis.** Top: relationships between median cLogP (**a**) or molecular mass (**b**) and promiscuity of 2,133 drugs and reference compounds in 200 assays from the Cerep BioPrint database. Promiscuity is the number of hits for each compound, in which >30% inhibition was found at a concentration of 10 µM. Points are coloured by median ring count and sized by either median molecular mass (281.3–582.9 Da) (**a**); or median cLog P (0.63–7.15) (**b**); promiscuity points with <4 compounds are omitted. The results of straight line fits ($n = 64$) are cLogP: slope 16.1, $r = 0.83$, $p = 1.4 \times 10^{-17}$; molecular mass: slope = 0.27, $r = 0.46$, $p = 0.0013$. The cLogP relationship appears sigmoidal and can be fitted to a logistic expression ($r = 0.84$). Bottom: the effect of ionization state, relationship between median log promiscuity and binned cLogP (**c**) and molecular mass (**d**). Labels are the number of compounds per point; points with <4 compounds are omitted. The total numbers of compounds (n) in each ionization class and median (mean) promiscuity values are: acids $n = 284$, 3 (5.1); bases $n = 813$, 23 (25.2); neutrals $n = 867$, 5 (8.4), zwitterions $n = 99$, 3 (4.7) and quaternary bases $n = 70$, 10 (19.1). The following equation describes promiscuity trends for the whole data set: Log Promiscuity = 0.075 cLogP – 0.71 A – 0.54 N – 0.47 Z + 1.00 ($n = 2133$; $r = 0.65$; $s = 0.42$; p for all variables and intercept < 10^{-25}; A, N and Z are indicator variables, set equal to 1 for acids, neutrals and zwitterions respectively). See also Supplementary information S5 (box) panels 3–6.

Pharmacophore
The ensemble of steric and electronic features that is necessary to ensure optimal interactions with a specific biological target structure and to trigger (or to block) its biological response.

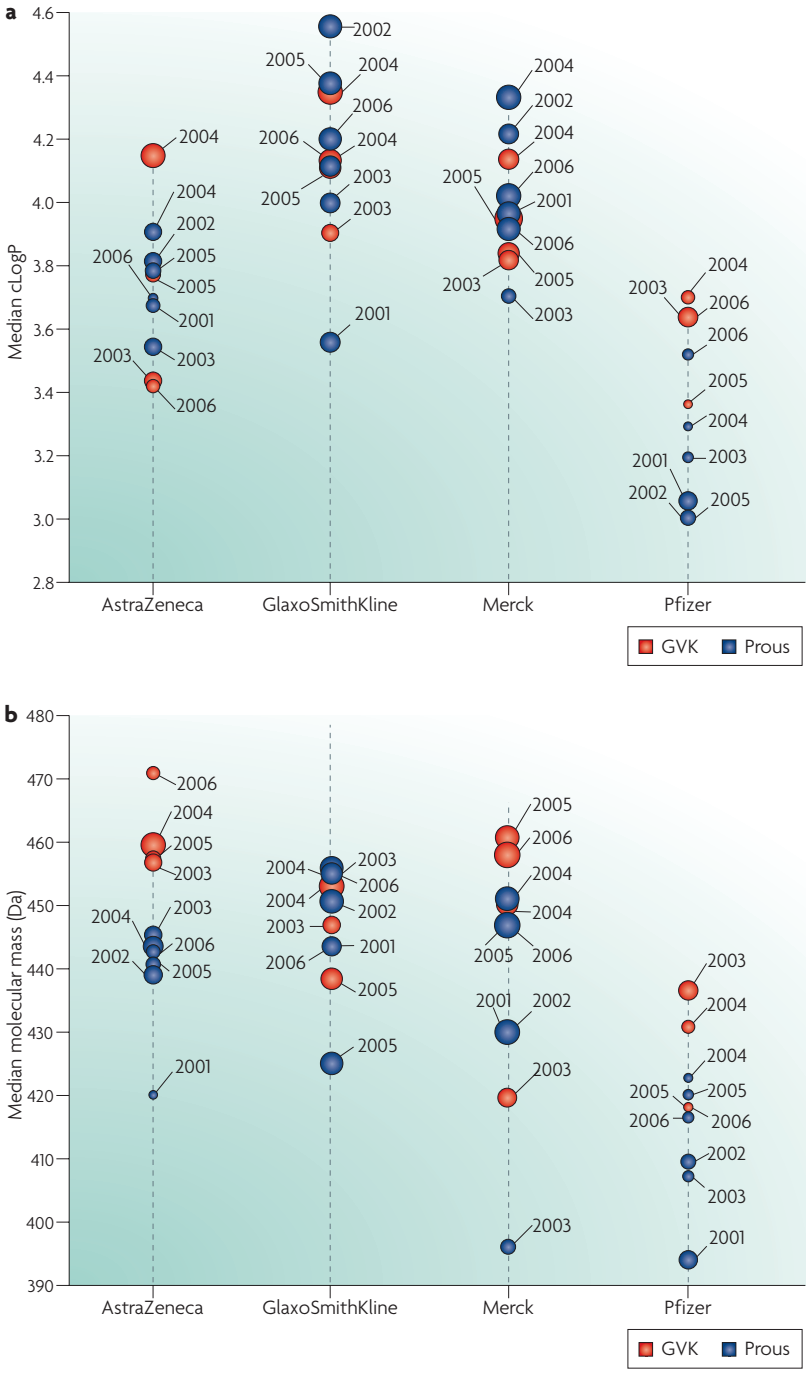

Figure 4 | **Trends in clogP and molecular mass in recently patented compounds from four pharmaceutical companies.** The median values of cLogP (**a**) or molecular mass (**b**) labelled by year obtained from WO patent databases (see BOX 1 for explanation of WO patent). Prous Science Integrity data covers 2001–2006 (blue) and GVK Bio patent data 2003–2006 (red). Points are sized by mean Lipinksi score (a value of 1 for each rule of 5 violation per compound; range here is 0.30–0.74). Aggregating the data and comparing companies gives the following statistically significant differences (for Student's t test, *p* values <0.05, see Supplementary information S5 (box), panels 7–12): Prous Science Integrity cLogP: AstraZeneca versus GlaxoSmithKline <0.0001; AstraZeneca versus Merck 0.0018; AstraZeneca versus Pfizer <0.0001; GlaxoSmithKline versus Merck 0.034; GlaxoSmithKline versus Pfizer <0.0001; Merck versus Pfizer <0.0001. GVK Bio cLogP: AstraZeneca versus GlaxoSmithKline 0.0004; GlaxoSmithKline versus Merck 0.036; GlaxoSmithKline versus Pfizer <0.0001; Merck versus Pfizer <0.0012. Prous Science Integrity and GVK Bio molecular mass: Pfizer versus the other companies <0.0001.

There is ample evidence that more extreme physical properties and therefore more complex molecules will have concomitantly increased predicted risks to developability — including bioavailability, permeability, solubility, synthesis and formulation — thereby decreasing chances of success through the development process[4,22].

Promiscuity

The fact that drug lipophilicity is changing less over time than other physical properties suggests that this is an especially important drug-like property, the control of which is important for ultimate success in drug development. This is not surprising: the role of LogP in influencing drug potency, pharmacokinetics and toxicity has been established for many years[22–26]. This property essentially reflects the key event of molecular desolvation in transfer from aqueous phases to cell membranes and to protein binding sites, which are mostly hydrophobic in nature. If lipophilicity is too high, there is an increased likelihood of binding to multiple targets and resultant pharmacologically based toxicology, as well as poor solubility and metabolic clearance.

The Cerep BioPrint database of drugs and reference compounds provides a unique opportunity to examine the role of physical properties in influencing drug promiscuity and side effects[27,28]. It is clear that overall promiscuity is predominantly controlled by lipophilicity and ionization state (FIG. 3). Bases and quaternary bases are notably more promiscuous than acids, neutral compounds or zwitterions (FIG. 3 and Supplementary information S3 (box)), but in all ionization classes, promiscuity correlates positively with cLogP (or ACDLogP, AZLogD7.4 and percentage of PSA, which are correlated with cLogP). Interestingly, the number of rings also shows a relationship with promiscuity, although this trend cannot be distinguished from lipophilicity.

The overall relationship between promiscuity and molecular mass is complex (FIG. 3b,d). Bases show optimal promiscuity in the 350–500-Da range and neutrals appear to be similar; however, with acids, there is a positive correlation between promiscuity and molecular mass (FIG. 3d). The results in FIG. 3 were corroborated by analysis of cross-screening data with AstraZeneca project compounds, which showed similar overall trends with ionization, cLogP and molecular mass. These results contrast with a published analysis[29] of high-throughput screening data, using 220 diverse targets and 75,000 compounds, in which promiscuity decreased proportionally with increasing molecular mass; perhaps the BioPrint data set is too small to reveal this trend. In addition to bulk lipophilicity, specific three-dimensional pharmacophoric and structural features including hydrogen bonding, polarizability and steric constraints, which are not fully represented by molecular mass, will drive the magnitude of binding affinity for protein active sites. Promiscuity is necessary for the action of some drugs, especially central nervous system agents, but it is clear that deliberate design of multiple-acting compounds at selected targets, an attractive therapeutic proposition, is a considerable

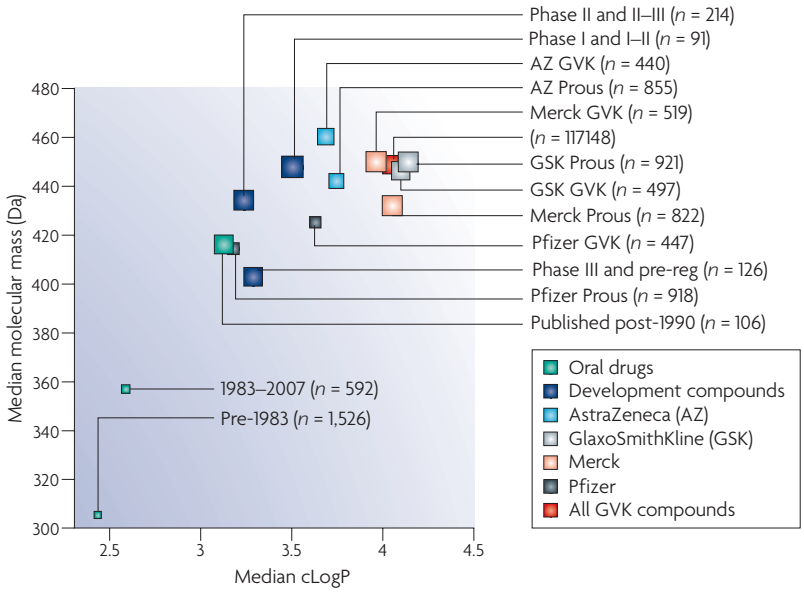

Figure 5 | **Overall trends in median cLogP and molecular mass in compounds from four pharmaceutical companies.** The figure compares historical oral drugs, compounds in development and in current patents from AstraZeneca, GlaxoSmithKline, Merck and Pfizer. Points are coloured and labelled by source, and sized by mean Lipinksi score (a value of 1 for each rule of 5 violation per compound; range here is 0.26–0.74). *n* = number of compounds, or number of patents for GVK Bio. For the aggregated Prous Science Integrity data over the period 2001–2006, there are upward property trends: the results of straight line fits of property versus year are molecular mass, slope 2.0 Da (both mean, $r = 0.61$ and median, $r = 0.62$, not significant) and cLogP slope 0.06 (mean, $r = 0.81$, $p = 0.05$) and 0.02 (median, $r = 0.31$, not significant).

physicochemical challenge[30]. Although promiscuity will be less important if selectivity for the desired target is high, the lesson from this analysis is that the risk of unwanted pharmacology increases with lipophilicity and is dependent on ionization class[66].

Current medicinal chemistry

The four companies with the most prolific contribution to the recent small-molecule drug discovery patent literature have differing year-by-year physical-property profiles among their patented compounds (FIG. 4). This might be expected given the timescales of discovery chemistry and project patent delivery, and overall project portfolios in a given year. However, looking at the property distributions over the whole period examined (2001–2007 for Prous Science Integrity data, 2003–2007 for GVK Bio data), there are significant differences between the companies. The rank order of cLogP and molecular mass values agree for both the Prous Science Integrity and GVK Bio patent data. For median cLogP, GlaxoSmithKline > Merck > AstraZeneca > Pfizer; for median molecular mass, GlaxoSmithKline = Merck = AstraZeneca > Pfizer. The overall correspondence between the two databases provides corroboration for the observed trends. The only differences are the AstraZeneca cLogP comparisons with Merck and Pfizer, which are statistically different in the Prous Science Integrity data (with more data points) but not in the GVK Bio patent data.

Taking the average property values for GVK Bio compounds aggregated from all companies confirms that the movement of chemical space in current chemistry to higher molecular mass and cLogP has progressed even further from historical oral drugs, recent oral drugs and development compounds. These comparisons are summarized in FIG. 5. The median patented compound has a cLogP of 4.1 and molecular mass of 450 Da, whereas the most recent oral drugs, discovered since 1990, have a median cLogP of 3.1 and molecular mass of 432 Da. Even in the 2001–2006 period, the upward trend in both cLogP and molecular mass is continuing in the aggregated Prous Science Integrity data.

One reason why physical properties may be increasing is that many of today's drug protein targets are different from those explored historically. It is clear that the four companies in this study are pursuing kinases and GPCRs responding to peptidic ligands as significant components of their portfolios (Supplementary information S4 (table)). Among drugs launched since 1983, these classes are a minority, with only 7 kinase inhibitors and 14 GPCR peptide inhibitors. Comparisons of target-class property profiles of current patents with oral drugs are summarized in FIG. 6, using the aggregated data from the four companies and the GPCR peptide and kinase data from the individual companies. Two main trends emerge from FIG. 6: first, the physical properties of most established drug classes are being inflated in current chemistry; second, the company differences seen in all patents persist in most cases among the kinase and GPCR peptide targets. The major compound target class in oral drugs approved since 1983 is the aminergic GPCR receptor (20% of the drugs), which is generally considered highly druggable as the endogenous ligands (serotonin, noradrenaline, histamine and acetylcholine) are small molecules. Even in this receptor class, both molecular mass and cLogP have significantly increased in current chemistry.

The final comparison between the companies is perhaps the most interesting. Chemokine GPCR receptors, responding to peptide ligands, have been among the more difficult target classes in terms of control of physical properties of antagonists. The chemokine (C-C motif) receptor 5 (CCR5) has been an area of considerable interest for treatment of HIV and rheumatoid arthritis, with the Pfizer CCR5 antagonist maraviroc[31] being the most advanced, having recently been approved for the treatment of HIV. The four companies arrived at the same chemical class of CCR5 antagonist, containing a common phenylpropylpiperidine pharmacophore (FIG. 7), as major components of their efforts on this target[31–34]. In this compound class there is a spread of 2.3 units in cLogP and 110 Da in molecular mass between companies (FIG. 7). The physical properties of the CCR5 phenylpropylpiperidine compounds are statistically different by company and this single class of compounds mirrors the overall company trends seen in FIG. 4. This single target analysis strongly suggests that differences in physical property inflation between companies are due to differences in local medicinal chemistry practice and the emphasis (or not) placed on the control of physical properties.

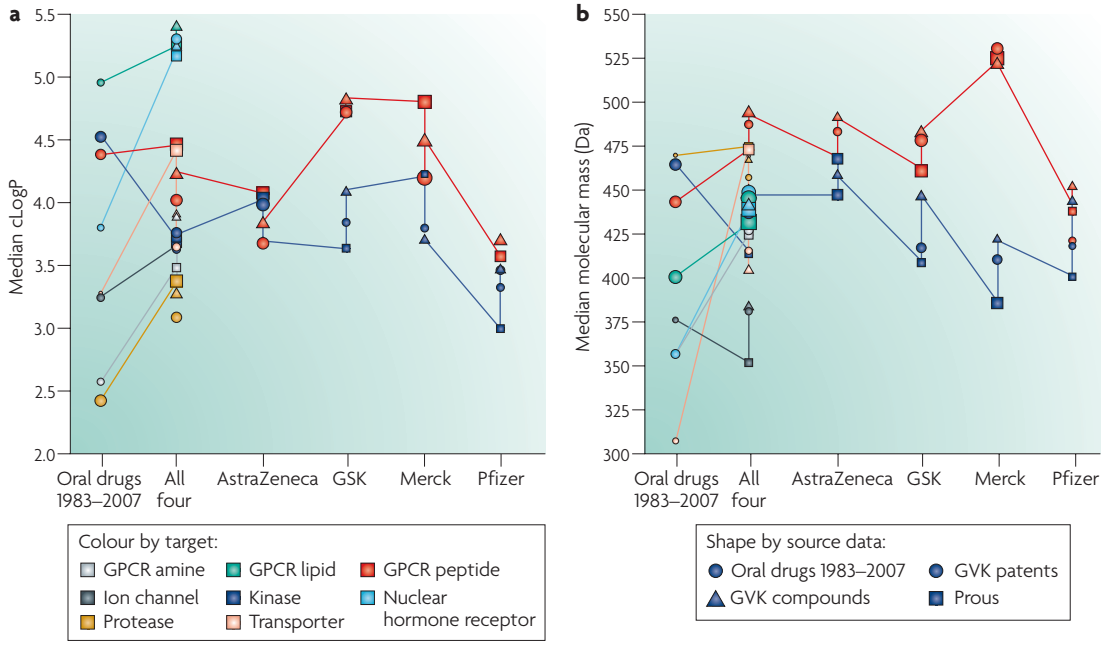

Figure 6 | **Target class trends in cLogP and molecular mass.** The median values of cLogP (**a**) and molecular mass (**b**) by drug target class, comparing oral drugs launched from 1983–2007 with individual companies (AstraZeneca, GlaxoSmithKline (GSK), Merck and Pfizer's G-protein-coupled (GPCR) peptide and kinase targets) and the aggregated company data (all targets). Points coloured by target class, sized by median molecular mass (307.8–531.0 Da) (**a**); or cLogP (2.44–5.41) (**b**); and shaped by source data. Comparison of companies was by Student's t tests (*p* values, ND = not different ($p > 0.05$)), see Supplementary information S5 (box), panels 13–24). For GPCR peptides, cLogP statistical values (Prous Science Integrity, GVK Bio patents) are: GlaxoSmithKline versus Pfizer, 0.012, <0.0001; GlaxoSmithKline versus AstraZeneca, 0.0032, <0.0001; GlaxoSmithKline vs Merck, ND, 0.025; Merck versus Pfizer 0.027, ND; Merck versus AstraZeneca 0.011, ND; AstraZeneca versus Pfizer, ND, ND. For molecular mass, statistical values (Prous Science Integrity, GVK Bio patents) are: GlaxoSmithKline versus Pfizer, 0.030, 0.0084; GlaxoSmithKline versus AstraZeneca, ND, ND; GlaxoSmithKline versus Merck, <0.0001, 0.0006; Merck versus Pfizer, <0.0001, <0.0001; Merck versus AstraZeneca, 0.0002, 0.0078; AstraZeneca versus Pfizer, 0.0019, 0.0007. For kinases, cLogP statistical values (Prous Science Integrity, GVK Bio patents) are: GlaxoSmithKline versus Pfizer, <0.0001, 0.0097; GlaxoSmithKline versus AstraZeneca, ND, ND; GlaxoSmithKline versus Merck, ND, ND; Merck versus Pfizer, <0.0001, ND; Merck versus AstraZeneca, ND, ND; AstraZeneca versus Pfizer, <0.0001, 0.0068. For molecular mass, statistical values (Prous Science Integrity, GVK Bio patents) are: GlaxoSmithKline versus Pfizer, ND, ND; GlaxoSmithKline versus AstraZeneca, 0.0005, <0.0001; GlaxoSmithKline versus Merck, ND, ND; Merck versus Pfizer, ND, ND; Merck versus AstraZeneca, 0.0002, 0.0002; AstraZeneca versus Pfizer, <0.0001, <0.0001.

Perspectives

Despite the apparent widespread acceptance of drug-like principles over the past decade, the trends in physical property inflation seen with recent oral drugs is continuing in current medicinal chemistry (FIG. 5). Why are chemists still synthesizing larger and more lipophilic compounds? With absorption, distribution, metabolism and excretion (ADME) optimization now a standard component of drug discovery programmes[22], persistent effort may pay off and result in compounds with good oral bioavailability that have physical properties in 'exception chemical space' versus oral drugs. Recent examples of high-molecular-mass oral drugs include HIV protease inhibitors and angiotensin II receptor antagonists; additionally, in a GlaxoSmithKline study, rat bioavailability was claimed to be unrelated to molecular mass[9]. Thus, pharmacokinetically acceptable chemical space can be found beyond the rule-of-five guidelines. Increasing lipophilicity will also tend to increase binding affinity[23–26], so the pursuit of lipophilic, large, potent and bioavailable compounds in current medicinal chemistry programmes probably explains the observed property inflation.

However, the most important challenge to the medicinal chemist today is not just obtaining high potency and good ADME, but also delivering candidate drugs that will not eventually fail owing to other compound-related properties, especially toxicity. The implications of working increasingly closer to the extremities of drug-like chemical space appear serious for overall productivity (FIG. 2) and promiscuity leading to increased risks of pharmacologically based toxicity (FIG. 3). Although more difficult target druggability[19,20] is playing a role in physical-property expansion, the effect of local medicinal chemical decision-making is significant as shown by the differences between companies (FIG. 4). In addition, changes in medicinal chemistry practice appear to be driving physical-property inflation in both newer and some older target classes (FIGS 6,7).

We suggest that compound lipophilicity, as estimated by cLogP, is the most important molecular property[23–26],

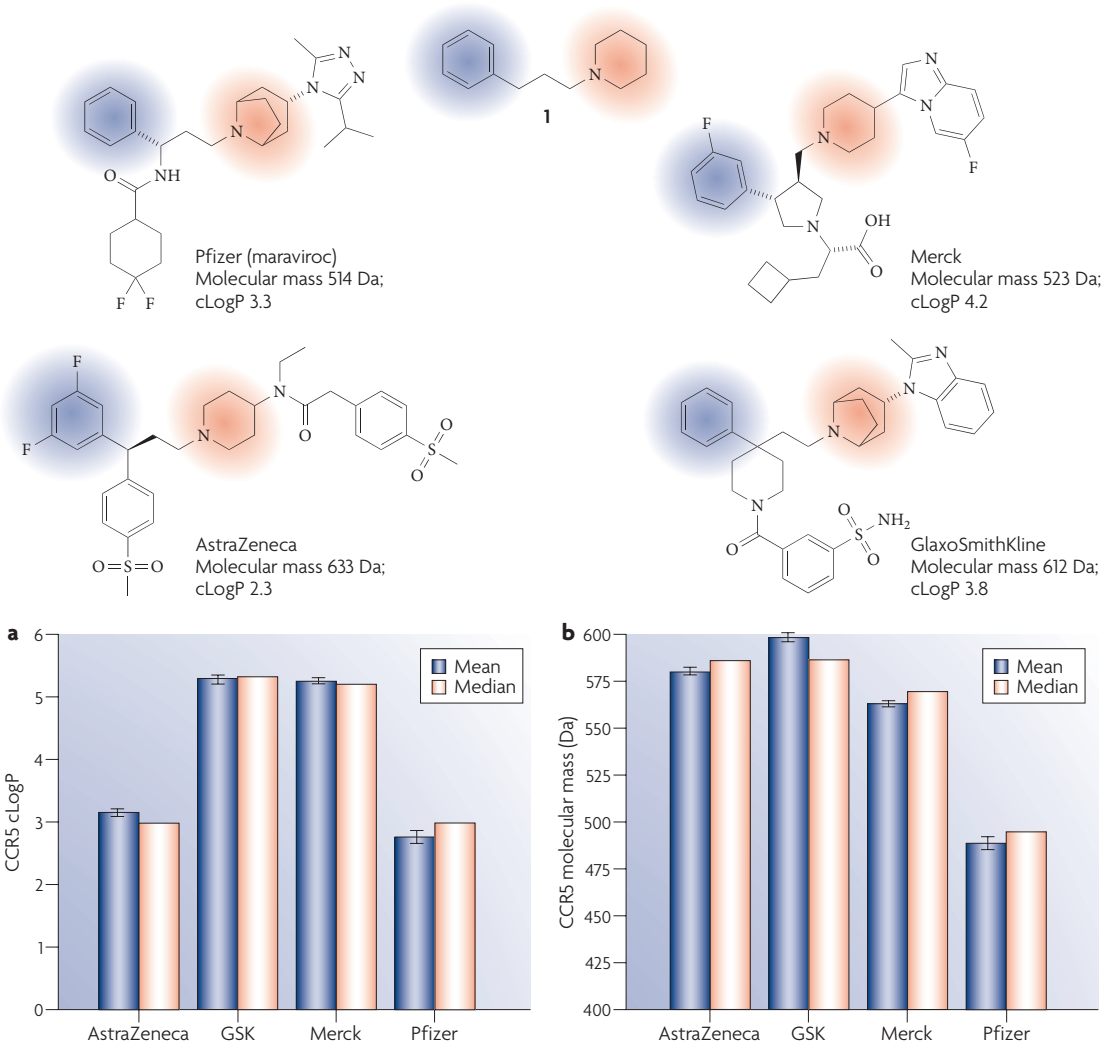

Figure 7 | **CCR5 as an example.** Representative structures and physical properties of published chemokine (C-C) receptor 5 (CCR5) antagonists based on a common phenylpropylpiperidine pharmacophore 1 from Pfizer[31], AstraZeneca[32], Merck[33], and GlaxoSmithKline (GSK)[34]. The mean and median values of cLogP (**a**) and molecular mass (**b**) abstracted from the GVK Bio database for all the companies patented CCR5 receptor antagonists containing pharmacophore 1. The numbers of phenylpropylpiperidine compounds and the proportion this represents of the companies' total CCR5 patented compounds are: AstraZeneca 1,069, 87%; GlaxoSmithKline 690, 46%; Merck 2,467, 82%; and Pfizer 309, 78%. Error bars are standard errors of the mean. Comparing companies gives the following statistically significant paired differences (Student's t test, p values <0.05, see Supplementary information S5 (box), panels 25,26): cLogP: all comparisons p <0.0001, except GlaxoSmithKline versus Merck, not significant. Molecular mass: all comparisons, p <0.0001.

as it is changing less over time in launched oral drugs than other properties (FIG. 1). Lipophilicity plays a dominant role in promoting binding to unwanted drug targets (FIG. 3). Lipophilic bases can cause cardiovascular toxicological effects by binding to the HERG (human ether-a-go-go-related potassium channel protein; also known as KCNH2) ion channel[35] and tissue toxicity by promoting cellular phospholipidosis[36]. But median current compounds from medicinal chemistry programmes conducted in leading pharmaceutical companies have cLogP values that have increased by ~1.5 log units relative to oral drugs launched from 1983–2007, and by ~1 log unit relative to the most recently discovered oral drugs. In addition, 30% of the patented compounds from the four companies in this study have cLogP values >5 (the rule-of-five cut-off). All these data suggest that the highly lipophilic compounds being made in many drug discovery programmes today carry increased risks of developmental attrition.

Ligand efficiency (LE) is an important new concept, which estimates the efficiency of a binding interaction with respect to the magnitude of ligand physical properties, most notably size (equation 1)[37]. We propose maximizing the minimally acceptable lipophilicity per unit of *in vitro* potency, or ligand-lipophilicity efficiency (LLE, equation 2)[37], as a more important objective for lead generation and optimization programmes.

$$LE = pIC_{50} \text{ (or } pK_i) \div \text{number of heavy atoms} \quad (1)$$

$$LLE = pIC_{50} \text{ (or } pK_i) - cLogP \text{ (or LogD)} \quad (2)$$

Phospholipidosis
Phospholipidosis is a lipid storage disorder in which excess phospholipids accumulate within cells. Drug-induced phospholipidosis occurs with many cationic amphiphilic drugs.

The average oral drug with cLogP ~2.5 and potency in the range ~1–10 nM suggests an LLE target of ~5–7 or greater. High *in vivo* potency has advantages: when the total dose in humans is low, adventitious compound-related toxicity is less of an issue[38]. In essence, the goal for optimization is to increase potency without increasing lipophilicity at the same time.

The differences between companies in this study are intriguing, and are important because the patented compound physicochemical profiles, and any potential resulting development risks, are likely to be reflected in the companies' pipelines. It is perhaps not surprising that the company with the lowest compound molecular mass and cLogP is Pfizer, in which the rule of five originated[1,2] and where property-based design has been practised for some time[22,39,40].

Identification of the lead compound, or chemical starting point, for a drug discovery programme has been highlighted recently as a critically important activity, reflected by lead generation strategies being widely implemented in the pharmaceutical industry[41–46]. In particular, the concepts of lead-likeness[47] and fragment-likeness[37,48] that followed drug-likeness propose that it is advantageous to use small, hydrophilic molecules to start drug discovery projects. This allows the processes of chemical optimization, which frequently increase physical properties[13,17,47,49,50], to work within drug-like space. Perhaps these strategic developments, which clearly emphasise 'small is beautiful,' have yet to be fully accepted or implemented? Medicinal chemists will additionally be influenced by other factors in their decision-making: their and their colleague's knowledge and experience[51], their intuition and creativity[52], their use of relevant predictive tools[39,40], their ability to handle vastly increased data flow[53] and even their prejudice[54]. Pursuit of different molecular targets does not seem to provide an obvious explanation for the differences between companies, as the overall trends are similar in current major target classes and in a single compound class (FIGS 6,7). Local cultures, policies (or lack of them), disease strategies, approaches to obtaining intellectual property, as well as specific project objectives and the drive for shorter drug discovery timelines, will also influence compound selection and optimization tactics.

Conclusion

Physicochemical properties in small-molecule drug discovery are completely under the control of medicinal chemists and can easily be calculated before chemical synthesis. With judicious selection of lead compounds and constant monitoring of physical properties (especially lipophilicity) during optimization, medicinal chemists have an opportunity to help alleviate the appalling attrition rates, estimated at 93–96% (REF. 55), in clinical drug development. A 5% improvement in attrition would double the output of new medicines: we suggest this might be achieved simply by lowering lipophilicity. It is time the medicinal chemistry community used its undoubted creative ability to better control physical properties, and to tackle the threat of compound-related toxicological attrition[67]. Failing to seek better compound quality would be irresponsible.

1. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.* **23**, 3–25 (1997).
2. Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov. Today Technol.* **1**, 337–341 (2004).
A perspective of the impact of the 'rule of five' in drug discovery.
3. Abraham, M. H., Chadha, H. S., Whiting, G. S. & Mitchell, R. C. Hydrogen bonding. 32. An analysis of water-octanol and water-alkane partitioning and the Δ log P parameter of Seiler. *J. Pharm. Sci.* **83**, 1085–1100 (1994).
4. Wenlock, M. C., Austin, R. P., Barton, P., Davis, A. M. & Leeson, P. D. A comparison of physiochemical property profiles of development and marketed oral drugs. *J. Med. Chem.* **46**, 1250–1256 (2003).
5. Vieth, M. *et al.* Characteristic physical properties and structural fragments of marketed oral drugs. *J. Med. Chem.* **47**, 224–232 (2004).
The physical property profiles of oral drugs are compared with topical and intravenous drugs.
6. Proudfoot, J. R. The evolution of synthetic oral drug properties. *Bioorg. Med. Chem. Lett.* **15**, 1087–1090 (2005).
7. Hou, T., Wang, J., Zhang, W. & Xu, X. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J. Chem. Inf. Model.* **47**, 208–218 (2007).
8. Linnankoski, J., Maekelae, J. M., Ranta, V.-P., Urtti, A. & Yliperttula, M. Computational prediction of oral drug absorption based on absorption rate constants in humans. *J. Med. Chem.* **49**, 3674–3681 (2006).
9. Veber, D. F. *et al.* Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **45**, 2615–2623 (2002).
10. Lu, J. J. *et al.* Influence of molecular flexibility and polar surface area metrics on oral bioavailability in the rat. *J. Med. Chem.* **47**, 6104–6107 (2004).
11. Martin, Y. C. A bioavailability score. *J. Med. Chem.* **48**, 3164–3170 (2005).
12. Hou, T., Wang, J., Zhang, W. & Xu, X. ADME evaluation in drug discovery. 6. can oral bioavailability in humans be effectively predicted by simple molecular property-based rules? *J. Chem. Inf. Model.* **47**, 460–463 (2007).
13. Morphy, R. The influence of target family and functional activity on the physicochemical properties of pre-clinical compounds. *J. Med. Chem.* **49**, 2969–2978 (2006).
A thorough study from the recent literature of medicinal chemical optimization practice across target classes, and consequent changes in physical-property profiles.
14. Oprea, T. I. Current trends in lead discovery: are we looking for the appropriate properties? *J. Comp Aided Mol. Design* **16**, 325–334 (2002).
15. Vieth, M. & Sutherland, J. J. Dependence of molecular properties on proteomic family for marketed oral drugs. *J. Med. Chem.* **49**, 3451–3453 (2006).
16. Paolini, G. V., Shapland, R. H. B., van Hoorn, W. P., Mason, J. S. & Hopkins, A. L. Global mapping of pharmacological space. *Nature Biotech.* **24**, 805–815 (2006).
Exemplifies the power of knowledge exploitation using large databases, providing insights into drug-like chemical space encompassing several hundred proteins.
17. Leeson, P. D. & Davis, A. M. Time-related differences in the physical property profiles of oral drugs. *J. Med. Chem.* **47**, 6338–6348 (2004).
18. Blake, J. F. Identification and evaluation of molecular properties related to preclinical optimization and clinical fate. *Med. Chem.* **1**, 649–655 (2005).
19. Cheng, A. C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nature Biotech.* **25**, 71–75 (2007).
Accurate 'druggability' predictions emerge from considerations of the predicted affinities of ligands for structurally characterized binding sites.
20. Hajduk, P. J., Huth, J. R. & Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **48**, 2518–2525 (2005).
21. Booth, B. & Zemmel, R. Prospects for productivity. *Nature Rev. Drug Discov.* **3**, 451–456 (2004).
22. van de Waterbeemd, H., Smith, D. A., Beaumont, K. & Walker, D. K. Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **44**, 1313–1333 (2001).
A key paper, laying down the essential principles and importance of physical-property optimization.
23. Hansch, C. Quantitative approach to biochemical structure–activity relationships. *Acc. Chem. Res.* **2**, 232–9 (1969).
24. Tute, M. S. Principles and practice of Hansch analysis. Guide to structure–activity correlation for the medicinal chemist. *Adv. Drug Res.* **6**, 1–77 (1971).
25. Kubinyi, H. Lipophilicity and biological activity. Drug transport and drug distribution in model systems and in biological systems. *Arzn. Forsch.* **29**, 1067–1080 (1979).
26. Cronin, M. T. D. The role of hydrophobicity in toxicity prediction. *Curr. Comp. Aided Drug Des.* **2**, 405–413 (2006).
27. Krejsa, C M. *et al.* Predicting ADME properties and side effects: the BioPrint approach. (http://www.cerep.fr/cerep/users/pages/collaborations/bioprint.asp). *Curr. Opin. Drug Disc. Dev.* **6**, 470–480 (2003).
28. Fliri, A. F., Loging, W. T., Thadeio, P. F. & Volkmann, R. A. Analysis of drug-induced effect patterns to link structure and side effects of medicines. *Nature Chem. Biol.* **1**, 389–397 (2005).
One of the first studies showing that *in vitro* binding profiles across multiple assays can be used to predict clinical side effects.
29. Hopkins, A. L., Mason, J. S. & Overington, J. P. Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.* **16**, 127–136 (2006).
30. Morphy, R. & Rankovic, Z. The physicochemical challenges of designing multiple ligands. *J. Med. Chem.* **49**, 4961–4970 (2006).

31. Price, D. A. *et al.* Overcoming HERG affinity in the discovery of the CCR5 antagonist maraviroc. *Bioorg. Med. Chem. Lett.* **16**, 4633–4637 (2006).
32. Cumming, J. G. *et al.* Modulators of the human CCR5 receptor. SAR of substituted 1-[3-(4-methanesulfonylphenyl)-3-phenylpropyl]-piperidinyl phenylacetamides. *Bioorg. Med. Chem. Lett.* **16**, 3533–3536 (2006).
33. Kim, D. *et al.* Potent 1,3,4-trisubstituted pyrrolidine CCR5 receptor antagonists: effects of fused heterocycles on antiviral activity and pharmacokinetic properties. *Bioorg. Med. Chem. Lett.* **15**, 2129–2134 (2005).
34. Kazmierski, W. M. *et al.* CCR5 Antagonists as Therapeutic Agents. World Patent WO2004054974 (2004).
35. Waring, M. J. & Johnstone, C. A quantitative assessment of hERG liability as a function of lipophilicity. *Bioorg. Med. Chem. Lett.* **17**, 1759–1764 (2007).
36. Baciu, M. *et al.* Degradative transport of cationic amphiphilic drugs across phospholipid bilayers. *Philos. Trans. A. Math. Phys. Eng. Sci.* **364**, 2597–2614 (2006).
37. Leach, A. R., Hann, M. M., Burrows, J. N. & Griffen, E. J. Fragment screening: an introduction. *Mol. Biosyst.* **2**, 429–446 (2006).
38. Uetrecht, J. Prediction of a new drug's potential to cause idiosyncratic reactions. *Curr. Opin. Drug Disc. Devel.* **4**, 55–59 (2001).
39. van de Waterbeemd, H. & Gifford, E. ADMET *in silico* modelling: towards prediction paradise? *Nature Rev. Drug Discov.* **2**, 192–204 (2003).
40. van de Waterbeemd, H. & Jones, B. C. Predicting oral absorption and bioavailability. *Prog. Med. Chem.* **41**, 1–59 (2003).
41. Lobell, M. *et al. In silico* ADMET traffic lights as a tool for the prioritization of HTS hits. *ChemMedChem* **1**, 1229–1236 (2006).
42. Wunberg, T. *et al.* Improving the hit-to-lead process: data-driven assessment of drug-like and lead-like screening hits. *Drug Discov. Today* **11**, 175–180 (2006).
43. Davis, A. M., Keeling, D. J., Steele, J., Tomkinson, N. P. & Tinker, A. C. Components of successful lead generation. *Curr. Top. Med. Chem.* **5**, 421–439 (2005).
44. Steinmeyer, A. The hit-to-lead process and Schering AG: strategic aspects. *ChemMedChem* **6**, 1–7, (2005).
45. Bleicher, K. H., Nettekoven, M., Peters, J.-U. & Wyler, R. Lead generation: sowing the seeds for future success. *Chimia* **58**, 588–600 (2004).
46. Jacoby, E. *et al.* Key aspects of the Novartis Compound Collection Enhancement Project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Top. Med. Chem.* **5**, 397–411 (2005).
47. Teague, S, J., Davis, A. M., Leeson, P. D. & Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem. Int. Ed.* **38**, 3743–3748 (1999).
The introduction of the 'lead-like' concept.
48. Makara, G. M. On sampling of fragment space. *J. Med. Chem.* **50**, 3214–3221 (2007).
49. Hann, M. M., Leach, A. R. & Harper, G. Molecular complexity and its impact on the probability of finding leads for drug discovery. *J. Chem. Inf. Comp. Sci.* **41**, 856–864 (2001).
A fundamental paper, showing that increased molecular complexity reduces the probability of binding.
50. Oprea, T. I., Davis, A. M., Teague, S. J. & Leeson, P. D. Is there a difference between leads and drugs? A historical perspective. *J. Chem. Inf. Comp. Sci.* **41**, 1308–1315 (2001).
51. Hopkins, A. L. & Polinsky, A. Knowledge and intelligence in drug design. *Annu. Rep. Med. Chem.* **41**, 425–437 (2006).
An interesting review discussing how medicinal chemists use their knowledge to make choices in drug discovery projects.
52. Lombardino, J. G. & Lowe, J. A. A guide to drug discovery: the role of the medicinal chemist in drug discovery — then and now. *Nature Rev. Drug Discov.* **3**, 853–862 (2004).
53. MacCoss, M. & Baillie, T. A. Organic chemistry in drug discovery. *Science* **303**, 1810–1813 (2004).
54. Leeson P. D., Davis, A. M. & Steele, J. Drug-like properties: guiding principles for design – or chemical prejudice? *Drug Discov. Today Technol.* **1**, 189–195 (2004).
55. Kola, I. & Landis, J. Can the pharmaceutical industry reduce attrition rates? *Nature Rev. Drug Discov.* **3**, 711–716 (2004).
56. Food and Drug Administration (FDA). Approved Drug Products with Therapeutic Equivalence Evaluations (FDA Orange Book). *FDA web site* [online], <http://www.fda.gov/cder/ob/default.htm> (2007).
57. Science Integrity®. <http://integrity.prous.com>
© Prous Science, a Thomson Scientific business; all rights reserved (2007).
58. Wood, A. *Annual Reports in Medicinal Chemistry Volume* **41** 503–522 (Academic Press, Burlington, 2006).
59. American Chemical Society. Chemical Abstract Database (CAS): Scifinder. *CAS web site* [online], <http://www.cas.org/SCIFINDER/> (2007) .
60. Merck Research Laboratories. The Merck Index. *Cambridge Soft web site* [online], <http://themerckindex.cambridgesoft.com/> (2007).
61. Gray, N. Changing Landscapes: A Special Report on the World's Top 50 Pharma Companies. *Pharmaceutical Executive web site* [online], <http://www.pharmexec.com/pharmexec/data/articlestandard//pharmexec/272006/354138/article.pdf> (2006)
62. GVK Bio. GVK Bio databases. *GVK Bio web site* [online], <http://www.gvkbio.com/informatics.html> (2007).
63. Beaumont, K., Schmid, E. & Smith, D. A. Oral delivery of G protein-coupled receptor modulators: an explanation for the observed class difference. *Bioorg. Med. Chem. Lett.* **15**, 3658–3664 (2005).
64. SAS Institute. JMP Statistical Software. *SAS Institute web site* [online], <http://www.jmp.com/software/> (2007)
65. TIBCO Software Inc. Spotfire DecisionSite product suite. *TIBCO Software web site* [online], <http://www.spotfire.com/products/decisionsite.cfm> (2007).
66. Azzaoui, K. *et al.* Modeling promiscuity based on *in vitro* safety pharmacology profiling data. *ChemMedChem* **2**, 874–880 (2007).
67. Bender, A. *et al.* Analysis of pharmacology data and the prediction of adverse drug reactions and off-target effects from chemical structure. *ChemMedChem* **2**, 861–873 (2007).

Acknowledgements
We gratefully acknowledge numerous fruitful discussions on physical properties with AstraZeneca colleagues S. Boyer, D. Cheshire, A. Davis, J. Dixon, D. Lathbury, J. Li, J.-E. Nyström, G. Pairaudeau, M. Rolf (who provided BioPrint data), J. Steele, S. Teague and M. Wenlock (who helped to assemble the oral drugs database). J. Proudfoot (Boehringer Ingelheim Pharmaceuticals) very kindly provided his oral drug database.

SUPPLEMENTARY INFORMATION
See online article: S1 (table) | S2 (box) | S3 (box) | S4 (table) | S5 (box)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF