versus nonoral marketed drugs, temporal patterns of development candidates versus marketed drugs, target family differences, and targeted simple analyses. Wenlock et al.[7] compared the mean and standard deviations of MW, log*P*, log$D_{7.4}$, HBD, HBA, and RB for orally administered clinical candidates from Phase I clinical trials to preregistration, as well as a set of 594 marketed oral drugs. The results showed that the mean molecular weight declined consistently as drug candidates advanced through the clinical trial process, going from 423 at Phase I to 337 in marketed oral drugs. Mean lipophilicity, as measured by ACD log*P*, was roughly constant (2.6 at to 2.5) but the discontinued development candidates at each phase had higher mean log*P* values (3.5 at Phase I, 3.5 at Phase II, 3.2 at Phase III). These differences were statistically significant and indicate there is an increased chance of clinical failure for high MW and/or log*P* compounds. Vieth et al.[8] examined the distributions of computed descriptors for 1,729 marketed drugs, including 1,193 orally administered drugs. They tabulated means, min/max, and different percentiles for 12 descriptors by six categories. One interesting and statistically significant difference was that injectable drugs have higher MW, greater polarity, lower lipophilicity, and are more flexible than oral drugs.

Two studies examined the changes in computed descriptors over time. For oral drugs launched prior to 1983, mean MW, HBA, RB, and number of rings are lower than for drugs launched during 1983–2002, whereas mean %PSA, Clog*P*, and HBD do not change significantly.[9] Similarly, Proudfoot[10] found that mean MW increased steadily from below 300 in 1950 to often above 400 in 1997 and that only seven drugs were marketed between 1937 and 1951 with MW $>$ 500 but that 32 drugs exceeding MW 500 were marketed 1983–1997. Lipophilicity did not increase. Increasing MW and steady lipophilicity causes an increase in polarity that would lower the probability of absorption. Also, Proudfoot notes that less than 5% of oral drugs have HBD $>$ 4, which may be related to their propensity for Phase II metabolism.

Studies of proteomic or target families show large differences in the distribution of computed descriptors between classes. Vieth and Sutherland[11] were able to assign a specific proteomic family to 642 of 1,210 marketed oral drugs. Mean descriptor values were not statistically different from overall oral drugs for drugs in the cytochrome P450, phosphodiesterase, kinase, and transporter families. Drugs targeting G-protein-coupled receptors (GPCRs) and proteases had significantly greater means for one or more of MW, Clog*P*, HBD, or HBA. Drugs targeting ion channels were significantly smaller than the overall distribution. Morphy[12] analyzed the computed property distributions of a literature and internal compound database at Organon containing data on 1,860 optimization projects. All target families showed increases in MW during optimization. Differences between families were due to differences in the properties of the leads. High property values were consistently observed for drugs targeting peptide GPCRs, integrin receptors, proteases, and transferases, whereas drugs targeting monoamine GPCRs, ion channels, oxidases, and transporters had lower property values. Antibacterial compounds have descriptor averages that are different from the average descriptor values reported for oral drug. Gram-positive antibacterials have average MW = 813, clog$D_{7.4}$ = −0.2, and PSA = 243. Gram-negative antibacterials have average MW = 414, clog$D_{7.4}$ = −2.8, and PSA = 165.[13]

Gleeson at GlaxoSmithKline has analyzed internal ADMET data on thousands of drug discovery compounds using only three simple descriptors: MW, log*P*, and ionization state. The analyses show general trends in line with common beliefs, but are imprecise.[14] Correcting computed log*P* for ionization by using computed log$D_{7.4}$ $<$ 5 as a cutoff was shown to pass approximately 50% of molecules with computed log*P* $>$ 5 in a Lipinski-type analysis.[15]

Overall, several useful concepts emerge from these analyses. Different targets and routes of administration may require biased property distributions and screening libraries for successful lead optimization. This could influence the eventual chances of project success and should be taken into account early by project leaders. Once more, optimization focused on potency has been shown again to lead to larger molecules which increases the potential for poor ADME properties. The extent of any ADME issues would of course depend on the structure of lead molecules. Larger, more lipophilic molecules historically have an increased rate of failure in the clinic. Finally, more rules using simple descriptors have been identified for culling molecules with poor ADME properties.

SOLUBILITY

Solubility is a property that depends on many complex factors. It is important to know the exact solid form of the molecule that was tested, the solvent used, and the performance characteristics of the experimental method. Molecules are commonly amorphous in form early in the research process, less pure, and are dissolved in dimethyl sulfoxide (DMSO) to create stock solutions for archival storage and high-throughput screening. These DMSO stocks are then diluted with buffer for activity and ADME in vitro screening assays. In later stage research, larger quantities of promising molecules are synthesized with the aim of producing a crystalline solid suitable for formulation and dosing in animal studies for pharmacology, pharmacokinetics, and toxicity. Salt forms, pH-dependent ionization, the existence of polymorphs with their varying solubilities, melting point of the crystal lattice, and the many available formulation solvents (water, polyethylene glycol, methylcellulose, organics, etc.) all influence measured solubility. Solubility can be measured with varying degrees of accuracy ranging from cheaper and faster, but less accurate and more variable, kinetic approaches using nephelometry or flow cytometry detection to “gold standard”