A Comparison of Physiochemical Property Profiles of Development and Marketed Oral Drugs

Mark C. Wenlock,* Rupert P. Austin, Patrick Barton, Andrew M. Davis, and Paul D. Leeson

Departments of Physical & Metabolic Science and Medicinal Chemistry, AstraZeneca R&D Charnwood, Bakewell Road, Loughborough, Leicestershire, LE11 5RH, United Kingdom

Received September 25, 2002

The process of drug discovery applies rigorous selection pressures. Marketed oral drugs will generally possess favorable physiochemical properties with respect to absorption, metabolism, distribution, and clearance. This paper describes a study in which the distributions of physiochemical properties of oral drugs in different phases of clinical development are compared to those already marketed. The aim is to identify the trends in physiochemical properties that favor a drug's successful passage through clinical development and on to the market. Two libraries were created, one of current development oral drugs and one of marketed oral drugs. Statistical analysis of the two showed that the mean molecular weight of orally administered drugs in development decreases on passing through each of the different clinical phases and gradually converges toward the mean molecular weight of marketed oral drugs. It is also clear that the most lipophilic compounds are being discontinued from development.

Introduction

The major focus of the research-based pharmaceutical industry is the discovery of safe, efficacious, new chemical entities (NCEs) for therapeutic targets. Drug discovery is a complex multivariate process, but the basic requirements for orally administered NCEs include intrinsic potency, oral bioavailability, no toxicological effects in humans, and a significant advantage over existing accepted therapies (if applicable). Although it is possible to predict, with varying accuracy, what a NCE will do when orally administered to humans, the full potential of a NCE is not known until it has been tested in clinical trials. Estimations of the attrition rate in development are as high as 90%.[1] The main reasons for attrition include lack of clinical efficacy, inappropriate pharmacokinetics, animal toxicity, adverse reactions in humans, commercial reasons,[2] formulation issues, etc.

Traditionally, the driving force of drug discovery has been the synthesis of novel structures that show increased potency.[3] Current research strategy increasingly recognizes the importance of pharmacokinetic information such as oral bioavailability and suitable half-life, and safety margins over unacceptable side effects/toxicity are often equally important as ligand specificity and selectivity. A compound must have a suitable pharmacokinetic profile such that an acceptable oral dose of the compound leads to a therapeutically effective concentration at its target site in the body for the required length of time. This pharmacokinetic profile is a complex function of properties such as dissolution, intestinal absorption, cellular permeability, binding to plasma proteins, turnover by metabolic enzymes and other clearance mechanisms, drug distribution and deposition. These properties are in turn influenced by the physical properties of the compound, i.e., molecular weight, lipophilicity, hydrogen bond donors and acceptors.[4−15] However, it is frequently found that the relationship between the physical properties of compounds and their in vitro potency at a particular biological target is much better understood than any relationship between the physical properties of compounds and their resulting pharmacokinetic profiles.

A pragmatic approach to gaining an understanding of the balance of physical properties that leads to a suitable pharmacokinetic profile for oral administration is to profile the properties of compounds that have some degree of success as orally administered drugs. Molecular weight was first profiled by Pidgeon et al.[16] and Lipinski et al.[17] later profiled a range of properties leading to the well-known "rules of 5". Lipinski's work was based on the results of profiling the calculated physical property data on a set of 2245 compounds chosen from the World Drug Index because it was presumed that they would have suitable solubility and permeability for oral administration (selection criteria required compounds to have United States Adopted Name (USAN) or International Nonproprietary Name (INN) name, identifying them as entering Phase II (PII) clinical trials). Compounds were removed that were likely to have poor clinical exposure. Further filtering removed polymers, peptides, quaternary ammonium, and phosphates. Lipinski found that approximately 90% of the remaining compounds had molecular weight less than 500, calculated log *P* less than 5, sum of hydrogen bond donors (as a sum of NH and OH) less than 5, and sum of hydrogen bond acceptors (as a sum of N and O) less than 10. Lipinski proposed that poor absorption and permeation are more likely when two or more of these limits are exceeded. Sakaeda et al.[18] reported similar exclusion criteria (molecular weight of more than 500, and *C* log *P* value of more than 5) that differentiated poorly absorbed drugs from good drug candidates by studying the physicochemical property profiles for 222 marketed oral drugs.

* Address correspondence to M. Wenlock. Tel.: +44 1509 644436. Fax: +44 1509 645576. E-Mail: mark.wenlock@astrazeneca.com.

10.1021/jm021053p CCC: $25.00 © 2003 American Chemical Society
Published on Web 03/05/2003

Table 1. Number of Compounds in Different Development Phases

development phase	no. of compds	percentage commonality of structures between development phase and marketed oral drugs data set/%
Phase I (PI)	92	1.1
Discontinued Phase I (DI)	31	0.0
Phase II (PII)	173	7.5
Discontinued Phase II (DII)	72	4.2
Phase III (PIII)	141	17.0
Discontinued Phase III (DIII)	48	6.3
Preregistration (Prereg)	66	16.7
Marketed oral drugs	594	

Two opposing arguments can be given when trying to understand the implication of Lipinski's analysis. One argument is that the process of drug development applies rigorous selection pressures such that compounds with an unsuitable balance of physical properties for oral administration are more likely to be unsuccessful. Properties of those surviving compounds do encode crucial information, not only with respect to solubility and permeability, but other drug metabolism and pharmacokinetic properties. The opposing argument is that a property profile of existing successful drugs only highlights a historical artifact. They provide information about the properties of the often structurally more simple drugs of the past aimed at more tractable biological targets, that are not representative of the greater complexity of modern drugs required to interact with more challenging biological targets, such as the modulation of protein−protein interactions. An ambiguity that further fuels this debate is that Lipinski's dataset contains compounds that did eventually make it to market as oral therapies along with compounds that eventually failed in PII or Phase III (PIII).

To further examine the relationships between physical properties and progress toward a successful oral drug, we have examined the property profiles through the development phases, of candidate drugs specifically targeted at oral administration, compared to a library composed of the current marketed oral drugs.

Methods

The web-based database R&D Insight[19] contains data on the development phase and status of compounds that have entered development since 1985. It is not uncommon to find a candidate drug in a different development phase depending on the country that clinical trials are being carried out in and/or route of administration being investigated. Hence R&D insight was searched for compounds intended for oral administration that were in development only in the USA up to January 2000. These compounds were compared with a database of 594 marketed (in the USA) oral drugs created from an analysis of the USA pharmacopoeia "The Physicians' Desk Reference 1999".[20] The number of compounds in each phase is shown in Table 1.

The following physical properties were calculated for the sets of development and marketed oral drugs: molecular weight, calculated logarithmic value of *n*-octanol/water partition coefficient (ACD LogP, version 4.0),[21] calculated logarithmic value of *n*-octanol/water distribution coefficient (ACD $LogD_{7.4}$, version 4.0),[21] number of hydrogen bond donors (H-bond donors), number of hydrogen bond acceptors (H-bond acceptors) and number of rotatable bonds (Cerius2 version 4.6).[22] For each of the calculated properties, standard statistical tests and calculations have been carried out to analyze how the properties vary through the stages of development. The t-tests (unpaired, 2-tailed, assuming unequal variance) have been carried out to determine the probability that the mean value of a particular physiochemical property for drugs in a particular development phase is significantly different from the mean value of that same property for marketed oral drugs. The use of these statistical procedures strictly requires that the data be approximately normally distributed and on a continuous scale. Normal probability plots show, for marketed oral drugs, that this is true for distributions of molecular weight, ACD LogP, and ACD $LogD_{7.4}$, but not for H-bond donors, H-bond acceptors, and rotatable bonds (Figure 1). The deviating tails on the normal probability plots for molecular weight, ACD LogP, and ACD $LogD_{7.4}$ only contain a small proportion of the total data sets (less than 8% in the worst case). Distributions for H-bond donors, H-bond acceptors, and rotatable bonds are skewed away from normality since the

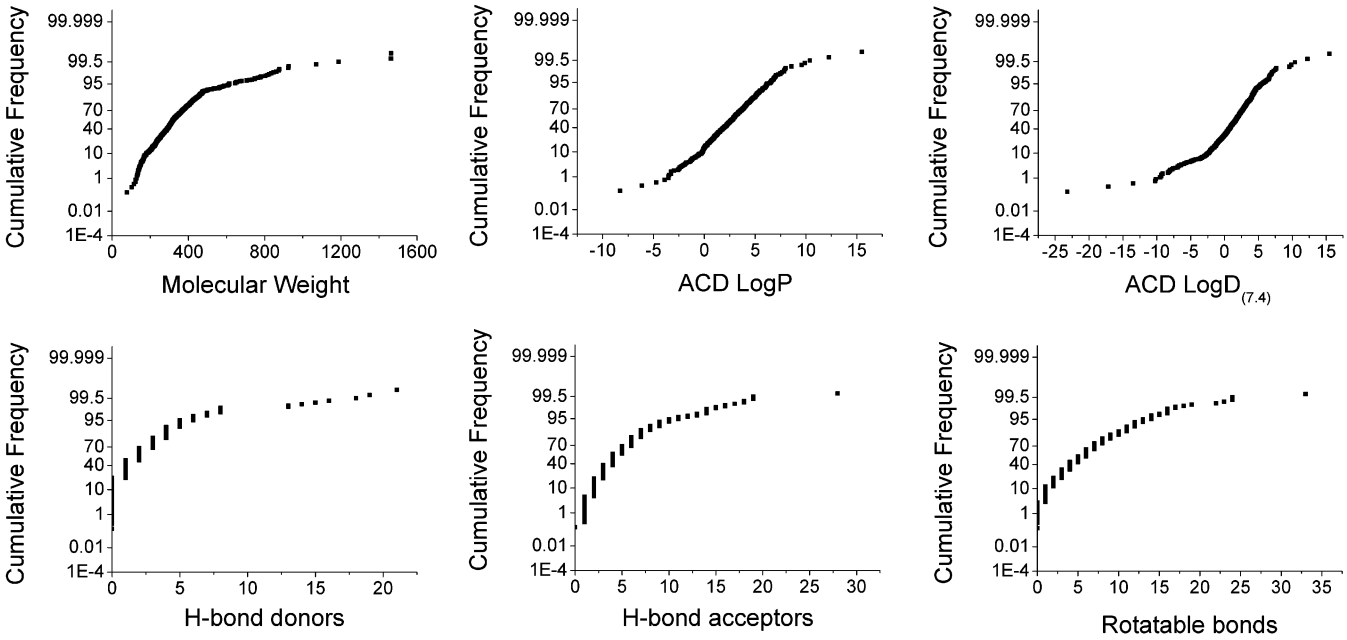

Figure 1. Normal probability plots of the distribution of molecular weight, ACD LogP, ACD $LogD_{7.4}$, H-bond donors, H-bond acceptors, and rotatable bonds for marketed oral drugs.

data contain a rather limited range of discrete values. The Mann−Whitney test (2-tailed), a nonparametric statistical procedure, was used to determine the probability that the distributions of H-bond donors, H-bond acceptors, and rotatable bonds for each phase are significantly different from the distributions shown by marketed oral drugs.[23] The value below which 90% of compounds lie is the classification used by Lipinski, which is simply the value of a particular property below which 90% of compounds in a particular phase are found.

Results

When comparing means of two data sets, t-tests and Mann−Whitney tests are used to determine significance levels. These statistical methods assume that the two data sets contain independent observations. Their application to this type of analysis is complicated by the fact that candidate drugs may be present in the same or a different development phase but for different indications. Even though certain compounds are common to multiple phases, each development phase data set has been filtered so that it contains only one example of each structure in that phase. However, some marketed oral drugs have been entered into new clinical trails to study efficacy for a different indication. The degree of commonality between the data set of a particular development phase and that of the marketed oral drugs data set is shown in Table 1. If a particular development phase data set includes common structures to that of the marketed oral drug data set then the use of these tests should be questioned as the two data sets do not contain independent observations. To overcome this, the common structures can be removed from the development phase data set but to do this would bias the observed physiochemical property limits of a particular development phase. In addition, it is possible to argue that a compound's successful passage through development to the market is independent of whether it is in a different stage of development, or has already been marketed, for a different indication since different indications will have differing criteria based primarily on existing therapies. For example, Pramipexole launched in the USA for Parkinson's disease but discontinued in PII in the USA for depression. This is partly due to the demand for new drugs to have better pharmacokinetics and safety profiles and partly to the increased commercial potential required by larger pharmaceutical companies. This dynamic change in the criteria for successful drugs raises an interesting, albeit academic, question of whether current oral drugs for a particular indication, if discovered today, would make it to the market for the same indication.

Calculated data and statistics for each of the six physical properties in each of the phases of development with and without (numbers in parentheses) common structures included are summarized in Table 2. Removing the common structures has very little effect on the results. Hence, reference to calculated data and statistics within the text and in Figures 2−8 is to the calculated data and statistics on the data sets with the common structures included.

Molecular Weight. Figure 2 clearly shows that the distribution of molecular weight for marketed oral drugs is very different from that of Phase I (PI), the latter being highly skewed to the right. The mean molecular weights for drugs in all phases of development are

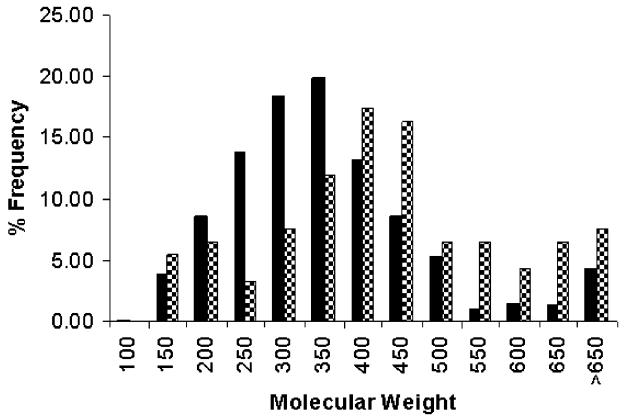

Figure 2. Distributions of molecular weight for marketed oral drugs (black) and development phase I oral drugs (checkered).

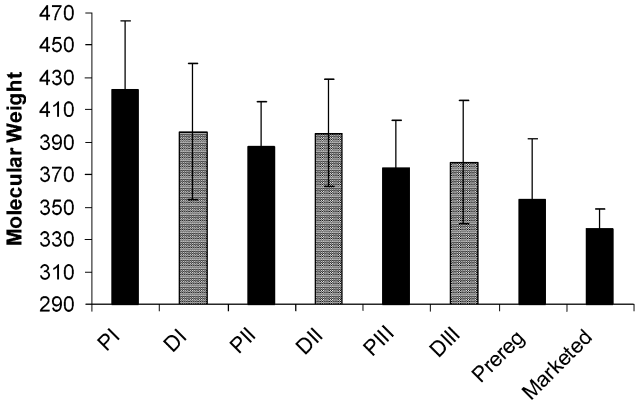

Figure 3. Mean molecular weight for drugs in different phases.

higher than that of marketed oral drugs. The t-tests show that this difference is significant at the 95% confidence level for PI to Discontinued Phase III (DIII) inclusive. Preregistration (Prereg) phase drugs, although having means greater than that shown by marketed oral drugs, are not significantly different at the 95% confidence level. Figure 3 is a bar chart displaying how the mean molecular weights vary throughout the various development phases. Included on the histogram are error bars showing the 95% confidence intervals of the means. This graph indicates that on passing from PI through to Prereg there is a gradual convergence toward a similar distribution of molecular weight as that shown by marketed oral drugs. There is a trend such that compounds discontinued from a particular phase have a higher molecular weight than compounds progressed into the next phase of development (mean molecular weight in Discontinued Phase I (DI) > mean molecular weight in PII, mean molecular weight in Discontinued Phase II (DII) > mean molecular weight in PIII and mean molecular weight in DIII > mean molecular weight in Prereg). This is consistent with Pidgeon's[16] suggestion that when the size of a compound approaches that of phospholipid molecules, the energetics of transporting such compounds across lipid bilayers becomes less favorable, thus reducing passive absorption. Larger molecules are also more likely to contain toxic pharmacophores or rapidly metabolized moieties.

ACD LogP. The mean ACD LogP of PI, PIII, and Prereg compounds are all very similar to that of

Table 2. Summary of Physiochemical Property Data for Different Phases of Development

physiochemical property	phase	mean	standard deviation	probability mean for a phase differs from that of marketed oral drugs/%	value below which 90% of compounds lie
molecular weight	PI	423 (423)	208 (209)	>99.9 (>99.9)	639 (639)
	DI	397	116	99.1	541
	PII	388 (395)	187 (191)	99.9 (>99.9)	533 (534)
	DII	396 (404)	143 (139)	99.9 (>99.9)	545 (559)
	PIII	374 (374)	179 (181)	97.7 (95.9)	500 (500)
	DIII	378 (384)	130 (132)	95.9 (97.4)	516 (516)
	Prereg	355 (349)	155 (151)	63.7 (44.2)	470 (458)
	Market	337	157		473
ACD LogP	PI	2.6 (2.6)	2.6 (2.6)	18.6 (7.8)	5.5 (5.5)
	DI	3.5	2.0	98.1	5.7
	PII	3.1 (3.1)	2.8 (2.8)	98.5 (97.5)	6.8 (6.8)
	DII	3.5 (3.6)	3.2 (3.1)	98.7 (98.9)	7.5 (8.2)
	PIII	2.5 (2.4)	2.5 (2.5)	19.9 (46.6)	5.6 (5.5)
	DIII	3.2 (3.3)	2.1 (2.2)	94.2 (96.2)	6.1 (6.1)
	Prereg	2.5 (2.3)	2.6 (2.7)	5.6 (53.6)	5.3 (5.0)
	Market	2.5	2.5		5.5
ACD $LogD_{7.4}$	PI	1.3 (1.2)	3.8 (3.7)	45.0 (37.8)	5.0 (4.8)
	DI	1.6	2.6	76.1	4.3
	PII	1.9 (1.9)	3.1 (3.2)	99.9 (99.7)	5.3 (5.4)
	DII	2.0 (2.0)	3.6 (3.6)	95.9 (96.2)	5.8 (5.8)
	PIII	0.8 (0.8)	3.4 (3.2)	42.3 (49.1)	4.6 (4.7)
	DIII	1.7 (1.8)	2.5 (2.5)	91.3 (94.2)	4.7 (4.7)
	Prereg	0.7 (0.3)	3.3 (3.3)	51.6 (84.5)	4.0 (3.3)
	Market	1.0	3.4		4.3
H-bond donors	PI	2.5 (2.5)	2.2 (2.2)	99.7 (99.6)	5 (5)
	DI	1.8	1.7	51.1	4
	PII	2.2 (2.2)	2.3 (2.3)	48.4 (63.0)	4 (4)
	DII	2.0 (2.0)	2.2 (2.2)	19.9 (30.8)	4 (5)
	PIII	2.3 (2.3)	2.2 (2.3)	94.9 (89.7)	5 (5)
	DIII	1.8 (1.8)	2.0 (2.1)	52.4 (71.5)	3 (3)
	Prereg	1.8 (1.9)	1.2 (1.2)	17.0 (32.2)	3 (3)
	Market	2.1	2.4		4
H-bond acceptors	PI	6.4 (6.4)	5.3 (5.3)	>99.9 (>99.9)	9 (9)
	DI	5.5	2.9	86.3	10
	PII	5.5 (5.5)	4.4 (4.5)	99.2 (99.7)	9 (9)
	DII	5.4 (5.5)	2.7 (2.7)	98.5 (99.5)	9 (9)
	PIII	5.5 (5.6)	4.5 (4.8)	99.2 (98.4)	8 (8)
	DIII	4.9 (5.0)	2.0 (2.1)	81.1 (84.4)	7 (7)
	Prereg	5.0 (5.0)	2.5 (2.4)	77.2 (80.1)	8 (8)
	Market	4.9	3.6		8
rotatable bonds	PI	7.8 (7.8)	6.7 (6.7)	99.5 (99.4)	14 (14)
	DI	7.1	5.4	77.8	14
	PII	6.8 (7.0)	5.5 (5.6)	97.8 (98.3)	12 (12)
	DII	7.9 (8.2)	5.6 (5.6)	99.9 (>99.9)	15 (15)
	PIII	7.3 (7.4)	5.4 (5.7)	>99.9 (>99.9)	13 (13)
	DIII	7.8 (8.0)	5.2 (5.3)	99.9 (>99.9)	11 (11)
	Prereg	5.8 (5.8)	4.0 (3.9)	1.9 (10.3)	12 (12)
	Market	5.9	4.5		11

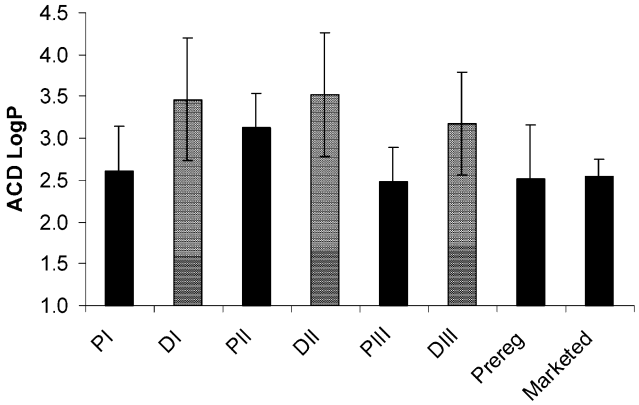

Figure 4. Mean ACD LogP for drugs in different phases.

marketed oral drugs (Figure 4). Included on the histogram are error bars showing the 95% confidence intervals of the means. Conversely, the mean ACD LogP of compounds in phases DI, PII, DII, and DIII are approximately 0.6 log units higher than marketed oral drugs, and this difference is significant at the 95% level for compounds in DI, PII, and DII, and significant at the 90% level for compounds in DIII. There is a trend such that compounds discontinued from a particular phase have a higher ACD LogP than compounds progressed into the next phase of development (mean ACD LogP in DI > mean ACD LogP in PII, mean ACD LogP in DII > mean ACD LogP in PIII and mean ACD LogP in DIII > mean ACD LogP in Prereg). This is consistent with the common finding that high lipophilicity frequently leads to compounds with rapid metabolic turnover,[4] and low solubility and poor absorption.[17]

ACD $LogD_{7.4}$. The mean ACD $LogD_{7.4}$ of PI, PIII, and Prereg compounds are all very similar to that of marketed oral drugs (Figure 5). Included on the histogram are error bars showing the 95% confidence inter-

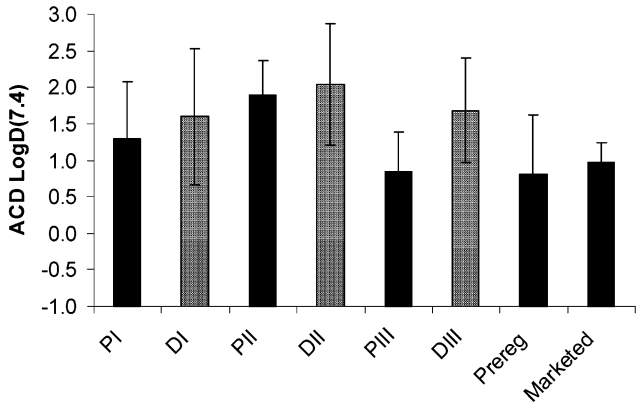

Figure 5. Mean ACD LogD (7.4) for drugs in different phases.

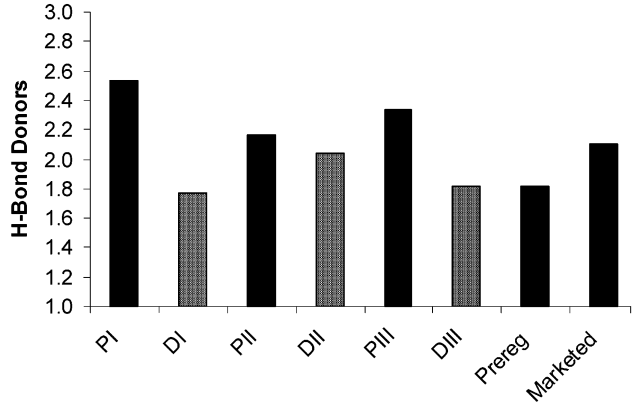

Figure 6. Mean number of H-bond donors for drugs in different phases.

vals of the means. However, the mean ACD $LogD_{7.4}$ of compounds in phases DI, PII, DII, and DIII compounds are approximately 0.6 log units higher than that shown by marketed oral drugs, with this difference being significant at the 95% level for compounds in PII and DII, and significant at the 90% level for compounds in DIII. The trend that compounds discontinued from a particular phase, have a higher ACD $LogD_{7.4}$ than compounds progressed into the next phase of development is present for DII to PIII and DIII to Prereg, but not shown between DI and PII. This implies that ACD $LogD_{7.4}$ is less important in influencing which compounds progress from PI to PII.

H-Bond Donors. The mean number of H-bond donors of compounds in marketed oral drugs is 2.1; in all the different development phases the mean number only varies between 1.8 for compounds in DI and Prereg to 2.5 for compounds in PI (Figure 6). The Mann−Whitney test analysis shows that only PI has a distribution of H-bond donors that is significantly different at the 95% level from the distribution shown by marketed oral drugs.

H-Bond Acceptors. The mean number of H-bond acceptors of compounds in marketed oral drugs is 4.9. In all the different development phases the mean number only varies between 4.9 for compounds in DIII to 6.4 for compounds in PI (Figure 7). The Mann−Whitney test analysis shows that PI, PII, DII, and PIII have a distribution of H-bond acceptors that is significantly different at the 95% level from the distribution shown by marketed oral drugs.

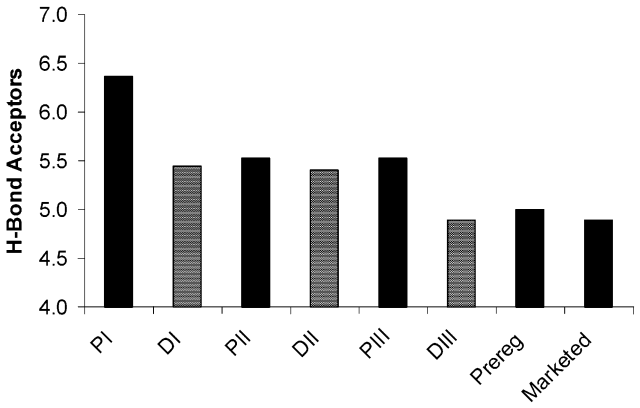

Figure 7. Mean number of H-bond acceptors for drugs in different phases.

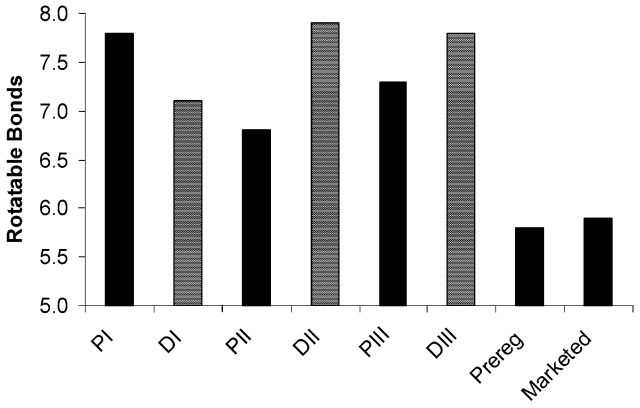

Figure 8. Mean number of rotatable bonds for drugs in different phases.

Rotatable Bonds. The mean number of rotatable bonds of compounds in marketed oral drugs is 5.9. In all the different development phases the mean number varies between 5.8 for compounds in Prereg to 7.9 for compounds in DII (Figure 8). The Mann−Whitney test analysis shows that PI, PII, DII, PIII, and DIII have a distribution of rotatable bonds that is significantly different at the 95% level from the distribution shown by marketed oral drugs. There is a trend such that compounds discontinued from a particular phase have a higher number of rotatable bonds than compounds progressed into the next phase of development (mean number of rotatable bonds in DI > mean number of rotatable bonds in PII, mean number of rotatable bonds in DII > mean number of rotatable bonds in PIII and mean number of rotatable bonds in DIII > mean number of rotatable bonds in Prereg). This trend is consistent with Veber's[24] suggestion that compounds with lower molecular flexibility, as measured by the number of rotatable bonds, tend to have better oral bioavailability.

Discussion

One of the clearest findings that can be drawn from this study is that the mean molecular weight of orally administered drugs in development decreases on passing through each of the phases and appears to gradually converge toward the mean molecular weight of marketed oral drugs data set. This is supported by the observation that the mean of the compounds discontinued from a particular phase is greater than the mean

of the compounds progressed to the next development phase. A similar trend, in which the mean of the compounds discontinued from a particular phase is greater than the mean of the compounds progressed to the next development phase, is apparent for the ACD LogP data and the ACD $LogD_{7.4}$ data (only in the latter stages of development), such that the most lipophilic compounds are being discontinued from development. This supports Lipinski's findings that there are limiting factors to the molecular weight and lipophilicity of a drug candidate that are reflected in the current physiochemical property profiles of the marketed oral drug data set. As molecular weight and lipophilicity are often positively correlated, this interdependence could in principle mask the relative importance of these properties. However, for the marketed oral drug data set molecular weight and lipophilicity show very little correlation (r^2 from linear regression = 0.04), and hence these properties show very little interdependence.

The marketed oral drug data set is made up of compounds each with an acceptable set of physiochemical properties that have successfully enabled it to overcome the obstacles of development for its desired therapeutic indication. As the route of administration is common to the whole data set, the physiochemical profiles of this data set reflects the range of properties that were once acceptable for progression to the market with respect to oral bioavailability. As the criteria for progression become inevitably more difficult, simply because for a drug candidate to progress beyond PIII it needs to show an advantage over standard accepted therapies (if applicable), the distributions of the physiochemical profiles for the chosen set of marketed oral drugs do not necessarily mirror the present day acceptable ranges for progression, and could reflect to a certain degree a historical artifact.

The very little difference in mean H-bond donor count between the various phases of development and marketed oral drugs suggests that, with respect to this property, compounds are already fairly well optimized from the early stages of development. There is evidence for development drugs having higher mean H-bond acceptor count than marketed oral drugs. The Mann−Whitney test shows that the distributions for a number of the development phases are different from that of marketed oral drugs (Table 2).

This analysis does show a trend in which the mean rotatable bond count of the compounds discontinued from a particular phase is greater than the mean of the compounds that exist in the next development phase, a pattern similar to that found with molecular weight. However, the extent of correlation between molecular weight and rotatable bond count (r^2 from linear regression for marketed oral drugs = 0.51) means that it is not possible to tell from this analysis which of these two properties is directly influencing successful progression. Molecular weight appears to show the clearest trends, but either molecular weight or rotatable bond count could be a surrogate for another correlated property or properties that exert a direct influence on successful progression to the market. For example, as highlighted earlier, increasing molecular weight will have a tendency to lead to lower permeability, lower solubility, increased number of metabolizable moieties, and toxic pharmacophores.

Table 3. Value below which 90% of Compounds Lie: Lipinski's Data Set versus Marketed Oral Drugs Data Set

	value below which 90% of compounds in data set lie	
physiochemical property	Lipinski's[17]	Marketed Oral Drugs
molecular weight	500	473
calculated log *P*	5 (*C* log *P*)	5.5 (ACD LogP)
calculated log $D_{7.4}$		4.3
H-bond donors	5	4
H-bond acceptors	10	7

In terms of the value of each property below which 90% of oral drugs lie, the values determined in this study from a database of marketed oral drugs are more stringent (see Table 3) than those determined by Lipinski for compounds that have reached at least PII clinical trials. This is most likely because Lipinski's dataset contains compounds that did eventually make it to market as oral therapies along with compounds that eventually failed in PII or PIII. Lipinski noted that only 1% of the compounds in his set lay outside his 90% limits for both molecular weight and lipophilicity (these are compounds with both molecular weight > 500 and calculated log *P* > 5), implying that these two properties, particularly in combination, are very important in determining the success of a compound intended as an orally administered drug. In this study, we have also found that the molecular weight and lipophilicity of candidate drugs are important in determining successful passage through the stages of development. The study shows that compounds with high molecular weight and/or lipophilicity are less likely to progress to the market.

Many of the compounds that make up the data sets for the varying development phases were presumably in early stage discovery projects back in the early 1990s. It is around this time that pharmaceutical companies started implementing high-throughput screening to aid drug discovery. It has been suggested that in the past, new drugs were discovered through subtle changes to lead structures which although gave rise to improved efficacy did not radically alter their physiochemical profile which were often already drug-like.[3,17] Although high-throughput in vitro screening allows for huge chemical libraries to be screened for potency in a short period of time, the issue with the process concerns the physiochemical property profile of the library being screened. If this library contains compounds with unfavorable properties for drug-like behavior, any potent hits from a screening campaign might struggle in subsequent optimization stages to simultaneously increase potency and improve pharmacokinetics by lowering molecular weight and/or lipophilicity. There is significant increase (86 Da) in the mean molecular weight of candidate drugs that have recently entered PI to those already on the market, but whether this increase has been influenced by the introduction of high-throughput in vitro screening is unclear. It has been suggested that high-throughput in vitro screening libraries need to be left shifted from the drug-like chemical space, particularly with respect to molecular weight and lipophilicity, and the lead-like paradigm has been introduced.[25,26] The findings of this study comple-

ment this recent concept of lead-like rather than drug-like molecules as preferred compounds for study in early discovery research.

Conclusion

The properties showing the clearest influence on the successful passage of a candidate drug through the different stages of development are molecular weight and lipophilicity. Statistical analysis shows that the mean molecular weight of orally administered drugs in development decreases on passing through each of the different clinical phases and gradually converges toward the mean molecular weight of marketed oral drugs. It is also clear that the most lipophilic compounds are being discontinued from development. This work supports Lipinski's findings that there are limiting factors to the molecular weight and lipophilicity of a candidate drug that are reflected in the current physiochemical property profiles of the marketed oral drug data set. In addition, this study suggests that these limiting values of physiochemical properties are not historical artifacts but are under physiological control.

References

(1) Prentis, R. A.; Lis, Y.; Walker, S. R. Pharmaceutical innovation by the seven UK-owned pharmaceutical companies (1964−1985). *Br. J. Clin. Pharm.* **1988**, *25*, 387−396.
(2) Venkatesh, S.; Lipper, R. A. Role of the development scientist in compound lead selection and optimization. *J. Pharm. Sci.* **2000**, *89*, 145−154.
(3) Curatolo, W. Physical chemical properties of oral drug candidates in the discovery and exploratory development settings. *Pharm. Sci. Technol. Today* **1998**, *1*, 387−393.
(4) van de Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. Property-based design: Optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313−1333.
(5) Riley, R. J.; Parker, A. J.; Trigg, S.; Manners, C. N. Development of a generalized, quantitative physicochemical model of CYP3A4 inhibition for use in early drug discovery. *Pharm. Res.* **2001**, *18*, 652−655.
(6) Lewis, D. F. V. On the recognition of mammalian microsomal cytochrome P450 substrates and their characteristics. Towards the prediction of human P450 substrate specificity and metabolism. *Biochem. Pharmacol.* **2000**, *60*, 293−306.
(7) Smith, D. A.; Ackland, M. J.; Jones, B. C. Properties of cytochrome P450 isoenzymes and their substrates part 2: Properties of cytochrome P450 substrates. *Drug Discovery Today* **1997**, *2*, 479−486.
(8) Rowley, M.; Kulagowski, J. J.; Watt, A. P.; Rathbone, D.; Stevenson, G. I.; Carling, R. W.; Baker, R.; Marshall, G. R.; Kemp, J. A.; Foster, A. C.; Grimwood, S.; Hargreaves, R.; Hurley, C.; Saywell, K. L.; Tricklebank, M. D.; Leeson, P. D. Effect of plasma protein binding on in vivo activity and brain penetration of glycine/NMDA receptor antagonists. *J. Med. Chem.* **1997**, *40*, 4053−4068.
(9) Barton, P.; Davis, A. M.; McCarthy, D. J.; Webborn, P. J. H. Drug-phospholipid interactions. 2. Predicting the sites of drug distribution using *n*-octanol/water and membrane/water distribution coefficients. *J. Pharm. Sci.* **1997**, *86*, 1034−1039.
(10) Chan, O. H.; Stewart, B. H. Physiochemical and drug-delivery considerations for oral drug bioavailability. *Drug Discovery Today* **1996**, *1*, 461−473.
(11) Palm, K.; Luthman, K.; Ungell, A.; Strandlund, G.; Artursson, P. Correlation of drug absorption with molecular surface properties. *J. Pharm. Sci.* **1996**, *85*, 32−39.
(12) Smith, D. A.; Jones, B. C.; Walker, D. K. Design of drugs involving the concepts and theories of drug metabolism and pharmacokinetics. *Med. Res. Rev.* **1996**, *16*, 243−266.
(13) van de Waterbeemd, H.; Camenisch, G.; Folkers, G.; Raevsky, O. A. Estimation of Caco-2 cell permeability using calculated molecular descriptors. *Quant. Struct.-Act. Relat.* **1996**, *15*, 480−490.
(14) Lázníček, M.; Lázníčková, A. The effect of lipophilicity on the protein binding and blood cell uptake of some acidic drugs. *J. Pharm. Biomed. Anal.* **1995**, *13*, 823−828.
(15) Austin, R. P.; Davis, A. M.; Manners, C. N. Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles. *J. Pharm. Sci.* **1995**, *84*, 1180−1183.
(16) Pidgeon, C.; Ong, S.; Liu, H.; Qiu, X.; Pidgeon, M.; Dantzig, A. H.; Munroe, J.; Hornback, W. J.; Kasher, J. S.; Glunz, L.; Szczerba, T. IAM chromatography: An in vitro screen for predicting drug membrane permeability. *J. Med. Chem.* **1995**, *38*, 590−594.
(17) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeny, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.
(18) Sakaeda, T.; Okamura, N.; Nagata, S.; Yagami, T.; Horinouchi, M.; Okumura, K.; Yamashita, F.; Hashida, M. Molecular and pharmacokinetic properties of 222 commercially available oral drugs in humans. *Biol. Pharm. Bull.* **2001**, *24*, 935−940.
(19) Adis International Ltd.
(20) *Physicians Desk Reference*, 53rd ed.; Medical Economics Company: Montvale, USA, 1999.
(21) Advanced Chemistry Development Inc., Toronto, Canada.
(22) Accelrys Inc., San Diego, California, USA.
(23) Daniel, W. W. *Applied Nonparametric Statistics*, 2nd ed.; Payne, M., Ed.; PWS−KENT Publishing Company: Boston, USA, 1990; pp 90−95.
(24) Veber, D. F.; Johnson, S. R.; Cheng, H.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.
(25) 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. Comput. Sci.* **2001**, *41*, 856−864.
(26) Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. The design of leadlike combinatorial libraries. *Angew. Chem.* **1999**, *111*, 3962−3967; Teague, S. J.; Davis, A. M.; Leeson, P. D.; Oprea, T. *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 3743−3748.

JM021053P