Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 1087–1090
Bioorganic & Medicinal Chemistry Letters

The evolution of synthetic oral drug properties

John R. Proudfoot*

Boehringer Ingelheim Pharmaceuticals Inc., 900 Ridgebury Road, PO Box 368, Ridgefield, CT 06877, USA

Received 26 August 2004; revised 30 November 2004; accepted 9 December 2004
Available online 12 January 2005

Abstract—An analysis of the properties of 1791 synthetic, oral drugs approved and/or marketed since 1937 demonstrates that the median molecular weight of oral drugs has increased substantially over the past 60 years. Fewer than 5% of approved/marketed oral drugs have more than 4-H bond donors and just 2% have MW > 500 and >3 H-bond donors.

Since the formulation of the Lipinski 'rule of five' which relates drug molecular weight (MW), Log *P*, and number of H-bond donors and H-bond acceptors to the potential for oral bioavailability,[1] a variety of insightful analyses of drug structure/property relationships has appeared.[2] One potentially interesting extension of these studies, which has been touched upon but not systematically examined, is the analysis of how these drug properties have evolved over time.[3] For example, Lipinski demonstrated an upward trend in MW for Merck and Pfizer clinical candidates over the period 1960–2000[4] while Vieth et al. showed that the mean MW and *C* Log *P* of oral drugs had not changed in the period 1982–2002.[2a] The first analysis is limited by the small sample size and focus on two drug discovery organizations: the second is limited by the narrow timeframe and by the use of FDA approval date to track changes with time. To develop such an analysis in a comprehensive manner it seemed worthwhile to expand on these studies as described below.

A list of approved or marketed drugs was generated by inspection of a variety of sources.[5] Two of these references[5g,h] were particularly useful in identifying drugs used only outside the USA and subsequently withdrawn or abandoned. The primary list derived from these sources contains the names of some 2900 substances. Refs. 5b, c and e also provided information on the route of administration.

With this information in hand, additional data was gathered and processed as follows. Patent priority date[6] or date of first literature disclosure was chosen to locate each drug and its properties in relation to time. The intention is to match, as closely as possible, each drug to its discovery date during the period covered, 1937–1997. Unmodified natural products were excluded from this analysis since these materials are generally not the product of drug design. Only drugs administered orally are included. Diagnostic agents and metal-containing drugs are excluded, as are enantiomers of drugs previously marketed as racemates. The final dataset contains 1791 oral approved or marketed drugs. Drug properties were calculated[7] on uncharged (except for quaternary ammonium and tertiary sulfonium) structures[8] stripped of counterions and solvates. Table 1 provides comparative statistics for this dataset and other recent similar datasets.

Table 1. Comparative statistics for different datasets

Parameter	This set	Lipinski[a]	Vieth[b]	Wenlock[c]
# oral drugs	1791	2245	1193	594
Mean MW	333		344	337
90th percentile	469	500	475	473
Mean Log *P*	2.5		2.3	2.5
90th percentile	4.8	5	5.2	5.5
Mean H donor	1.5		1.8	2.1
90th percentile	3	5	3	4
Mean H acceptor	5.1		5.5	4.9
90th percentile	9	10	9	7

[a] See Ref. 1.
[b] See Ref. 2a.
[c] See Ref. 2b.

Figures 1 and 2 display plots of MW, *A* Log *P*,[9] H-bond donors and N + O count (the original surrogate for H-bond acceptor ability[1]) versus Year (disclosure date). For the MW and *A* Log *P* plots datapoints for all 1791

* Tel.: +1 203 798 5107; fax: +1 203 791 6072; e-mail: jproudfo@rdg.boehringer-ingelheim.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.12.024

drugs are displayed along with the median, mean, maximum and minimum values connected by year.[10] For the H donor and N + O count plots, the number of records associated with each datapoint is shown. In each plot, the Lipinski rule value is highlighted in orange. At the right hand side of each plot the number of drugs falling within *Y*-axis property range or having that *Y*-axis property value is displayed, along with the percentage falling outside the particular Lipinski rule.

The plot of MW over time (Fig. 1, top) shows mean and median MW increasing from generally below 300 in the period 1937–1950 to higher values often above 400 from 1980 onwards. Given the lengthy timeframe from

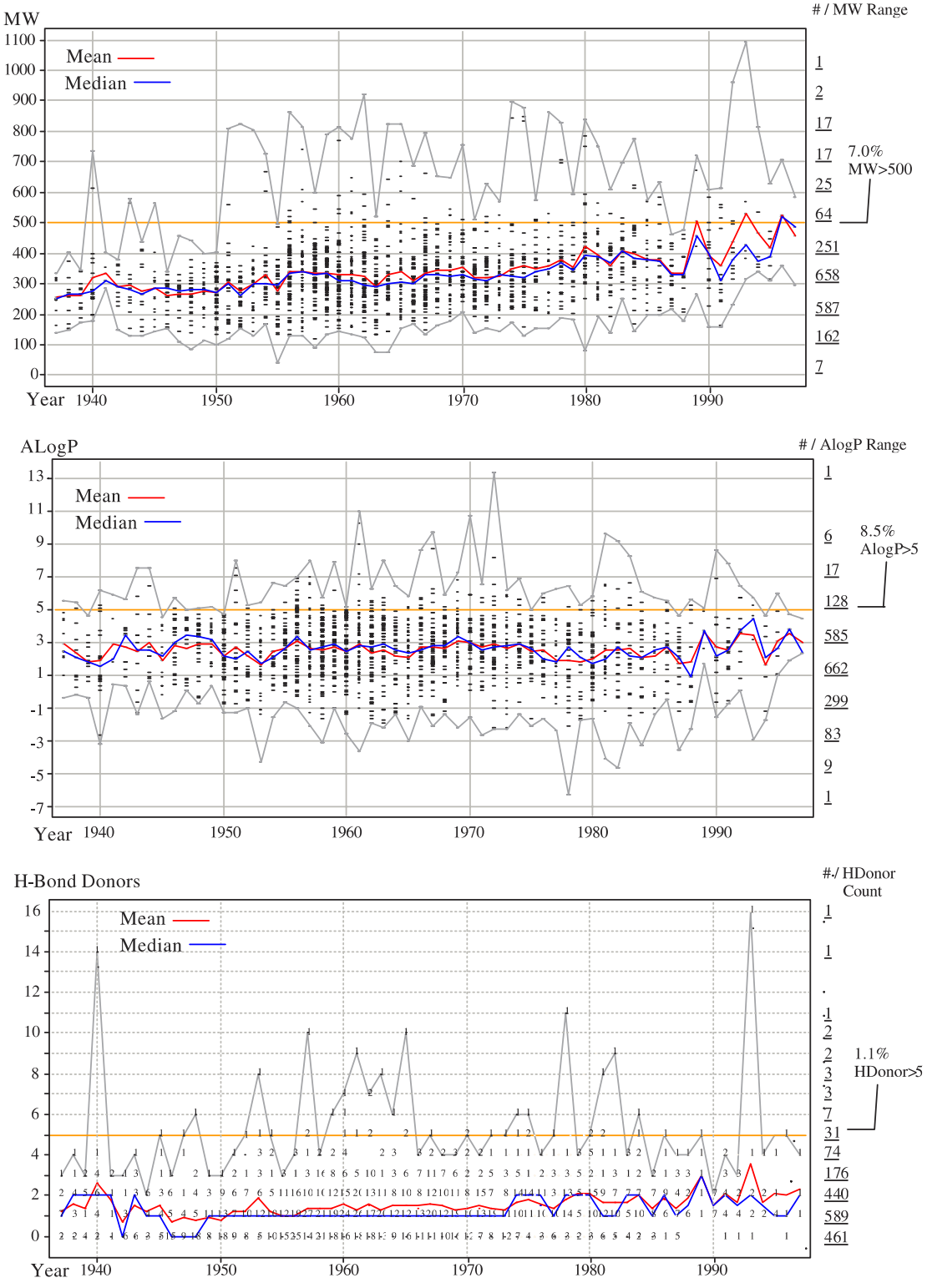

Figure 1. Drug properties versus time.

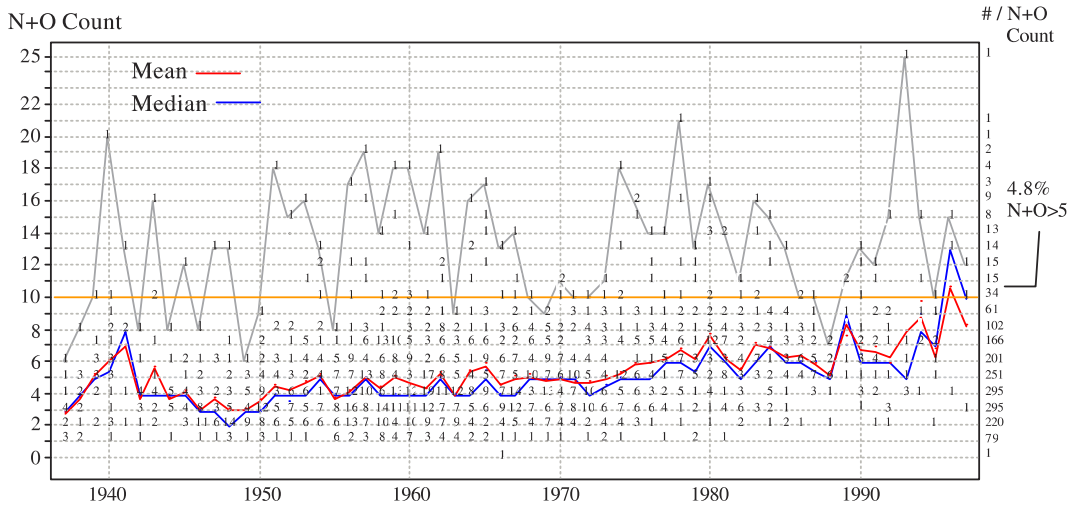

Figure 2. N + O count versus time.

discovery to approval, datasets for 1990 and beyond are incomplete and subject to future update, but the trend is upwards. For all the records, 7% have MW > 500 compared to 11% in Lipinski's dataset.[1] The plot of $A\log P$ over time (Fig. 1, centre) shows no upward or downward trend. 8.5% of records have $A\log P > 5$ compared to 10% with $C\,\mathrm{Log\,P} > 5$ in Lipinski's dataset. Only 5.2% of the drugs have $A\,\mathrm{Log}\,P < -1$.

The H-bond donor plot provided the biggest surprise, not because of a significant upward trend over time but because of the small number (1.1%) of structures violating the 5 H-bond donor rule. Since different computational tools calculate H-bond donors in different ways, we evaluated this property using additional protocols as shown in Table 2. Calculation of the H-bond donors at pH 7.4, to provide a more physiologically relevant count, gave essentially the same result (albeit with individual molecules distributed differently). Cerius2 returned a larger number of molecules with >5 H-bond donors, but still only 2.2% of the total. Given that the improvement of physical properties, most frequently solubility, is an ongoing challenge in most lead optimization programs it is enlightening to note that the indiscriminate appendage of multiple H-bond donors is not a likely path to success. Additionally, Oprea has commented that H bond donors are frequently involved in phase 2 metabolism and a low H donor count may be desirable for both absorption and adequate exposure.[11]

Table 2. Number of H donors from different protocols

H donors	Scitegic	Scitegic @pH 7.4	Cerius2
0	461	382	461
1	589	732	510
2	440	431	384
3	176	146	213
4	74	57	139
5	31	22	44
6	7	8	19
7	3	2	7
8	3	5	7
9	2	1	1
10	2	3	3
11	1		
12		1	1
13			
14	1		1
15			
16	1	1	
17			
18			1
>3	7.0%	5.6%	12.5%
>4	2.8%	2.4%	4.7%
>5	1.1%	1.2%	2.2%

For H-bond acceptors (Fig. 2, N + O count) an upward trend is apparent from the mid-1970s onward with more substantial increases in the 1990s. However, as noted above, the dataset for this later time period is necessarily incomplete. 4.8% of records have N + O count >10 compared to 12% in the Lipinski dataset.

Finally, we looked at the distribution of molecules violating more than one Lipinski rule. Figures 3 and 4

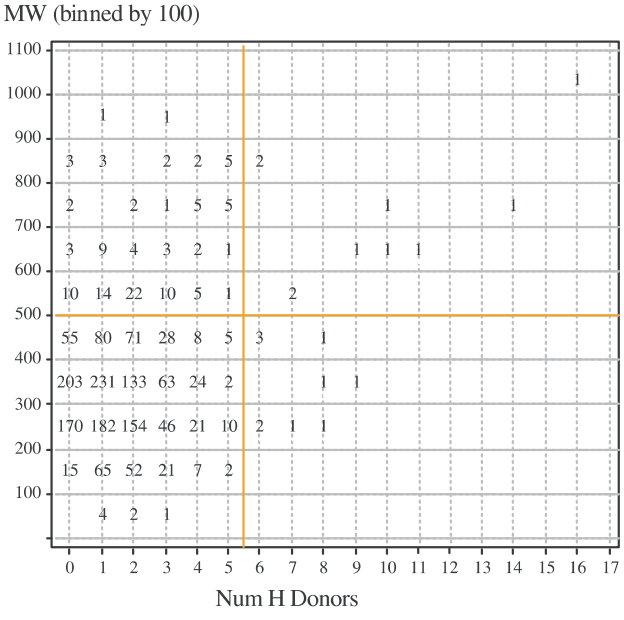

Figure 3.

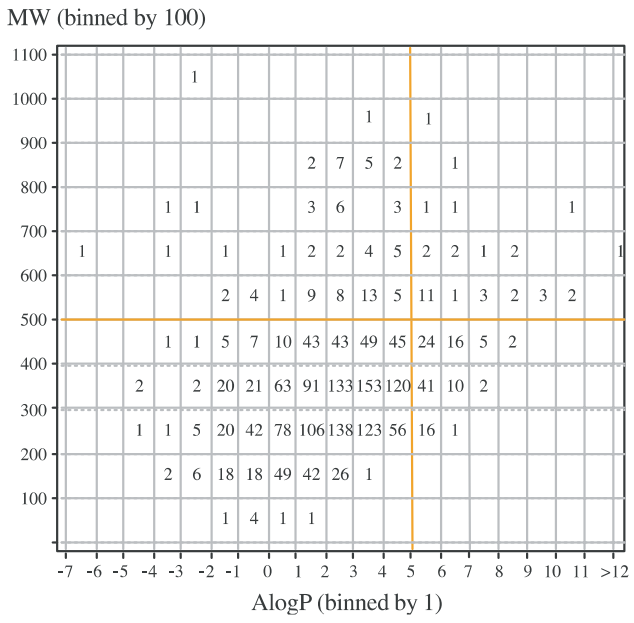

Figure 4.

depict the distribution of H-bond donors and (binned) $A\log P$ relative to MW binned in 100 amu increments.

Lipinski noted that 4% of his dataset violated MW and H-bond donor rules. Only 0.6% of this dataset of oral marketed drugs have MW > 500 and H-bond donors > 5 and just 2% have MW > 500 and H-bond donors > 3. Of these structures, 2% have MW > 500 and $A\log P > 5$ comparable to the 1% violating both the MW and $C\log P$ rules.

During the period 1937–1997 the most notable change in drug properties has been a steady increase in mean and median MW. Only seven marketed drugs with MW > 500 were designed in the 15 year period 1937–1951. The comparable period 1983–1997 has so far given thirty two. Mean and median lipophilicity have not changed in the 60 year period examined. Fewer than 5% of oral marketed drugs have more than 4 H-bond donors and an almost insignificant number exceed both the MW and H-bond donor rules.

Acknowledgements

I thank Dr. Jim Stevenson and Qiang Zhang for providing the computed drug properties.

References and notes

1. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliver. Rev.* **1997**, *23*, 3.
2. (a) Vieth, M.; Siegel, M. G.; Higgs, R. E.; Watson, I. A.; Robertson, D. H.; Savin, K. A.; Durst, G. L.; Hipskind, P. A. *J. Med. Chem.* **2004**, *47*, 224; (b) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. *J. Med. Chem.* **2003**, *46*, 1250; (c) Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.
3. Leeson, P. D. and Davis, A. M. *J. Med. Chem.* **2004**, *47* 6338.
4. Lipinski, C. A. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235.
5. (a) Annual Reports in Medicinal Chemistry, volumes 19–38. Information on worldwide launches, 1983–2002; (b) *FDA Approved Drugs Since 1938*, FOI Services, Inc. Electronic database of FDA approved drugs 1938–1998; (c) *Drugs Available Abroad, A Guide to Therapeutic Drugs Available and Approved outside the U.S.*; Schlesser, J. L., Ed.; Derwent: London, 1991, Information on non-USA approved products; (d) *Index Nominum International Drug Directory*; Medpharm GmbH Scientific: Stuttgart, 2004, Information on worldwide products; (e) *Martindale the Complete Drug Reference*; Sweetman, S. C., Ed.; Pharmaceutical Press: London, 2002, Information on worldwide products; (f) Graul, A. I. *Drug News Perspect.* **2004**, *17*, 43, Drugs reaching the market, 2003; (g) Sneader, Walter. In *Drug Prototypes and Their Exploitation*; John Wiley & Sons: Chichester, 1996. Information on marketed drugs; (h) *Consolidated List of Products Whose Consumption and/or Sale Have Been Banned, Withdrawn, Severely restricted or Not Approved by Governments*. United Nations: New York, 2004. Information on worldwide withdrawn drugs; (i) http://www.fda.gov/cder/rdmt/ FDA approvals 2003, 2004.
6. Most of the patent priority dates were obtained at the European Patent Office website: http://ep.espacenet.com/. Where this resource proved insufficient, the priority or patent publication date from Scifinder® was used.
7. Properties were calculated using Scitegic Pipeline Pilot® and Cerius2® software. The data shown is generated with Scitegic protocols. A table of the calculated properties is available from the author upon request.
8. Chemical structures were obtained from ChemIDPlus (http://chem.sis.nlm.nih.gov/chemidplus/), National Library of Medicine website, or Scifinder®.
9. $A\log P$ calculations are used here. The values obtained are comparable to the $C\log P$ data presented in Refs. 1 and 2b Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. *J. Phys. Chem. A* **1998**, *102*, 3762.
10. Statistics were generated in Excel®, plots in Spotfire®.
11. Oprea, T. I. *J. Comput. Aid. Mol. Des.* **2000**, *14*, 251.