LogD Contributions of Substituents Commonly Used in Medicinal Chemistry

Matthew L. Landry and James J. Crawford*

Genentech, Inc., 1 DNA Way, South San Francisco, California 94080, United States

Supporting Information

ABSTRACT: The importance of physicochemical property calculation and measurement is well-established in drug discovery. In particular, lipophilicity predictions play a central role in target design and prioritization. While significant progress has been made in our ability to calculate both $\log P$ and $\log D$, the quality of these predictions is limited by the size and diversity of the underlying data set. Access to diverse data sets and advanced models is often limited to large organizations or consortia, and they are not available to many students and practitioners of medicinal chemistry. A molecular matched pair analysis of median $\Delta\log D_{7.4}$ contributions for substituents commonly used in drug discovery programs at Genentech is reported. The results of this $\Delta\log D$ analysis are compiled into a single table, which we anticipate will be of use to practicing medicinal chemists.

KEYWORDS: lipophilicity, matched molecular pairs, $\log D$, $\Delta\log D$

Lipophilicity is a critical parameter in drug design and development.¹ Canonized in Lipinski's rule of five, trends in lipophilicity often correlate with the absorption, distribution, metabolism, excretion, and toxicity (ADMET) profile of drugs, while high lipophilicity has been shown to correlate with attrition in the development phase.^{1–4} For these reasons, measurement of the partition coefficient ($\log P$) or its pH-dependent variant, the distribution coefficient ($\log D$), is now routine in the process of drug discovery.⁵

The ability to predict the $\log D$ of molecules is central to the process of target design, evaluation, and prioritization within a medicinal chemistry program. This notion is accentuated by data that suggest the optimal lipophilicity for drugs exists within a narrow range of $\log P = 1–3$.¹ Pioneering work by Hansch, Leo, and Rekker demonstrated that the total lipophilicity of a molecule can be calculated through summation of the lipophilicity of its constitutive fragments.^{6–12} Building on this work, a variety of methods now exist for the prediction of $\log P$; a limitation of these methods is that they possess a “reasonable degree of error which is often systematic”.^{1,13,14} Advances in computing (e.g., machine learning) and use of large internal data sets have led to the creation of tailored models for lipophilicity prediction; however, these models are not broadly distributed or accessible to those at institutions with limited historical data.¹⁵ Owing to these factors, we surmised that a list of experimentally determined $\log D$ values for common molecular fragments based on a large, diverse, and pharmaceutically relevant set of compounds would provide value to medicinal chemists as a reference to aid in target design. While not allowing for an exact calculation of $\log D$, such a list provides a generalized

sense of lipophilic contributions of commonly employed substituents in the context of drug-like molecules.

Our approach to establish the lipophilicity contributions of functional groups leveraged Genentech $\log D$ data collected using the shake-flask method at pH 7.4 and an in-house molecular matched pairs (MMPs) application.^{5,16} MMP analysis has found use in the assessment of quantitative structure–activity relationships, including in studying the lipophilicity of substituents, although previous lipophilicity reports have been limited in scope.^{17–22} Using the MMP algorithm, lists of MMPs that differ by a defined functional group were generated, allowing for statistical analysis of the contribution to $\log D$ of that functional group. As $\Delta\log D$ for an MMP is expected to be highly dependent on the environment of substitution (e.g., in the case of H-to-Me substitutions, an MMP could represent quite diverse transformations, such as acid-to-ester or phenyl-to-tolyl changes), analysis was conducted at two different radii. Here, radius is defined as the number of bonds from the point of substitution in an MMP that must be preserved as a minimal, shared substructure when data from the analysis is compiled. The analysis presented here is based on radius values of zero and three and a substructure of phenyl. Radius = 0 means that the functional group in question can have any attachment to the matched pair parent molecule and no matched pairs would be excluded during analysis. A substructure of phenyl and radius = 3 means that the substitution is occurring specifically on a 1,4-disubstituted phenyl ring shared between the MMPs. While its use is well-

Received: October 25, 2019

Accepted: December 11, 2019

Published: December 11, 2019

Table 1. Lipophilic Contributions of Common Functional Groups^a

R	median $\Delta\log D_{\text{radius}=0}$	median $\Delta\log D_{\text{radius}=3}$	π ($\Delta\log P$) ^b	R	median $\Delta\log D_{\text{radius}=0}$	median $\Delta\log D_{\text{radius}=3}$	π ($\Delta\log P$) ^b
Me	0.30 (8458)	0.50 (112)	0.31/0.58	OH	-0.70 (1224)	-0.50 (45)	-1.12/0.07
Et	0.72 (634)	1.00 (15)	1.02	OMe	-0.05 (1579)	0.00 (135)	-0.02/0.71
<i>n</i> -Pr	1.05 (66)	2.00 (1)	1.45/1.55	OEt	0.40 (100)	0.70 (7)	0.03/0.51
<i>n</i> -Bu	1.42 (18)	1.60 (3)	2.13	<i>On</i> -Pr	1.00 (10)	1.05 (2)	1.05
<i>i</i> -Pr	1.10 (419)	1.10 (26)	1.22/1.53	<i>Oi</i> -Pr	0.85 (76)	0.80 (11)	0.36/0.92
<i>t</i> -Bu	1.30 (88)	1.40 (21)	1.98	OCF ₃	1.00 (138)	0.95 (40)	1.04
CFH ₂	0.20 (35)	0.05 (1)	–	OCF ₂ H	0.50 (88)	0.35 (20)	0.31
CF ₂ H	0.40 (260)	0.22 (6)	–	OAc	-0.10 (12)	–	-0.64/-0.27
CF ₃	0.90 (899)	0.80 (64)	0.10/1.07	NO ₂	0.02 (28)	0.40 (4)	-0.85/0.54
Ph	1.40 (570)	1.60 (19)	1.80/2.15	NH ₂	-0.30 (818)	-1.40 (9)	-1.42/-0.07 -4.19 ^c
CN	-0.28 (1092)	-0.30 (98)	-0.84/-0.15		-0.10 (120)	-0.80 (1)	-0.67/-0.47
F	0.20 (4249)	0.10 (230)	-0.17/0.14	NMe ₂	-0.30 (258)	-0.10 (21)	-0.30/0.69
Cl	0.60 (1782)	0.70 (198)	0.36/0.71		0.45 (90)	0.60 (4)	–
Br	0.90 (356)	0.80 (43)	0.60/0.92		0.80 (71)	0.80 (9)	0.85
I	1.10 (35)	1.30 (4)	1.00/1.12		-0.20 (278)	-0.50 (28)	–
0.90 (586)	1.05 (12)	1.14		-1.90 (53)	-2.00 (4)	–
1.30 (88)	1.80 (1)	–		-0.60 (139)	-0.40 (9)	-0.97/0.04
-0.60 (560)	-0.90 (24)	-1.03/-0.04		-1.00 (76)	-0.90 (12)	-1.18
-0.10 (103)	-0.30 (10)	-0.78	SMe	0.45 (30)	0.50 (3)	0.61
-2.33 (72)	-2.60 (23)	-1.04 -4.09 ^c		-0.90 (192)	-1.15 (38)	-1.63
-1.10 (57)	-2.40 (5)	-0.16/-0.15		-1.10 (21)	0.00 (3)	–
-2.98 (106)	-2.90 (9)	-0.32/-0.02 -4.67/-4.36 ^c		-0.52 (16)	-0.57 (6)	–
-1.10 (269)	-1.30 (22)	-1.71/-1.07		-0.15 (20)	-0.25 (6)	-0.78/0.03
-0.50 (140)	-1.00 (9)	-1.27
-0.98 (148)	-1.15 (26)	-1.51
-1.90 (62)	–	–				

^a $\log D$ measured by the shake-flask method at pH 7.4. Values in parentheses denote the number of MMPs for a given functional group at a given radius. ^b π -Values reported by Hansch and Leo.^{6–12} In cases where multiple data values were reported, they are expressed in the form minimum/maximum reported value. ^c π -Values for the ionized form of the functional group.

established in the literature, the term ‘radius’ is often not well-defined. Thus, a schematic illustration of the concept of radius in the context of this work is provided in the [Supporting Information](#) (Figure S1).

The median value for $\Delta\log D_{7.4}$ was chosen to limit the effects of experimental outliers, and the number of MMPs for a given functional group is provided in parentheses (Table 1). All $\Delta\log D_{7.4}$ values are compared with previously reported π -values ($\Delta\log P$ values for substituted benzene MMPs) from Hansch and Leo, when available; in cases where multiple values are listed, the minimum and maximum values are provided.^{7–9} Importantly, whereas the π -values provide data on a single, aryl-substituted MMP, this $\Delta\log D_{7.4}$ analysis provides a median measure of multiple MMPs of diverse, pharmaceutically relevant structure. In general, good agreement is observed between $\Delta\log D_{7.4}$ and π -values for non-ionizable groups. Analysis of aliphatic fragments reveals that the lipophilic contributions of these functional groups is greater in MMPs with a radius = 3 than with a radius = 0. This trend is reversed for fluoroalkanes, which show reduced lipophilicity in MMPs with a radius = 3. For ionizable motifs, as expected the median $\Delta\log D_{7.4}$ falls between the $\Delta\log P$ of the un-ionized and completely ionized functional group, thus offering insight in to the median pK_a values of these functional groups. For example, this analysis would suggest that primary amines and carboxylic acids in this library have median pK_a values of approximately 8.7 and 4.5, respectively. As might be expected, there are fewer matched pairs at radius = 3 because, in this analysis, this would require matched pairs of only 1,4-disubstituted phenyl groups. Of perhaps greater note, although the $\log P$ data reported by Hansch and Leo are extensive, no values are provided for some fragments that are now frequently observed in drug-like compounds, perhaps reflective of the development of synthetic methods in the intervening period.^{7–9}

Phenyl substitution is one of the most lipophilic changes studied in this analysis, and bioisosteric replacement of this motif is a routine means of reducing compound lipophilicity.^{23,24} Common phenyl bioisosteres include nitrogen-containing analogues of benzene such as pyridines and diazines. Analysis of the lipophilic contributions of these heterocycles was performed and found to span nearly a 2-log-unit range, including 3-pyridazine, which decreases lipophilicity by -0.80 units (Figure 1). These data provide a quantitative assessment of the lipophilicity of common phenyl replacements.

Cognizant of the observed standard deviation (see Table S1) for various substituents as well as variability in data measurements and its possible effect on our analyses, we sought to more closely scrutinize the underlying experimental data for median $\Delta\log D_{7.4}$. To this end, a detailed case study on trifluoromethyl MMPs was conducted which examined the most extreme outliers within this data set (compounds 1–6, Figure 2). The global maximum and minimum for $\Delta\log D_{7.4}$ of a trifluoromethyl substituent were $+2.0$ and -2.2 , belonging to the MMPs of 3 to 4 and 5 to 6, respectively (Figure 2, top). Given that these experimental measures are subject to error, a reassessment of these outliers was undertaken and produced data in closer agreement with the median $\Delta\log D_{7.4}$ ($+1.1$, in both cases). The measured $\log P$ of these compounds is in good agreement with the remeasured $\log D_{7.4}$ for compounds 4 and 6, while for compound 2, the discrepancy between $\log D_{7.4}$ and $\log P$ can be explained through a change in the measured pK_a of

Figure 1. Median $\Delta\log D$ values measured at pH 7.4 for common heterocycles. Analysis performed between MMPs (heterocycle- vs H-substituted) with radius = 0, and the number of MMPs is shown in parentheses.

its aryl sulfonamide ($\Delta pK_a = -0.7$ for 1 to 2). Postulating that a source of error in these initial $\log D$ measurements might arise from poor solubility, the kinetic solubility of each of these compounds was tested. The observed error in initial $\log D$ measurement correlated with decreased solubility, in line with the necessity for compound solubility in $\log D$ determination by the shake-flask method.²⁵ This observation led to two questions: (1) could variability in $\log D$ measurements be mitigated by filtering out compounds with limited solubility, and (2) would filtering by solubility affect median $\Delta\log D_{7.4}$? To this end, functional group data sets were filtered with kinetic solubility limits of >25 and $>100 \mu\text{M}$, resulting in a decrease in the standard deviation for $\Delta\log D_{7.4}$ without a notable effect on median $\Delta\log D_{7.4}$ (Figure 2, bottom). This effect was widely observed among the functional groups tested and demonstrates that (1) limited solubility is correlated with error in $\log D_{7.4}$ measurement and (2) removing outliers through filtering by solubility tightens standard deviation while not meaningfully changing median $\Delta\log D_{7.4}$. Further data and additional statistics for $\Delta\log D$ contributions of all substituents shown in Table 1 are provided in the [Supporting Information](#) (Table S1).

Although filtering by compound solubility removes many outliers, there still exists a range of $\Delta\log D_{7.4}$ values for functional groups. This provides insight in to the breadth of lipophilic contribution that a functional group may have in the context of forming an MMP. Such a range is partially reflective of the unique environment of substitution between two MMPs. For example, some substitutions may lead to changes in ionization state of nearby functionality, alterations in molecular conformation, masking of polarity, and minimization or maximization of dipoles. All of these outcomes can drastically impact the contribution of a given functional substitution. Notwithstanding these ranges of contribution, the size and diversity of the data set used in this analysis provide a generalized metric for the lipophilicity of standard functional groups. Importantly, as lipophilicity will always be context-dependent, a range of contributions will always exist for a given functional group; however, median measures of $\Delta\log D$ taken on a diverse array of scaffolds can provide a guide for target design.

The application of MMP analysis to Genentech’s database of compounds has allowed for the generation of a table of

Figure 2. Top: Lipophilicity, solubility, and acidity data for outlier CF₃-MMPs at radius = 3. Bottom: Median and mean values for ΔlogD among CF₃-MMPs at radius = 0 (left) and radius = 3 (right), along with statistical data for the same data sets filtered by kinetic solubility thresholds of >25 and >100 μM. Kin. Sol. = kinetic solubility, SD = standard deviation, Count = number of MMPs, and rad = radius.

lipophilicity values that we anticipate will provide value to the medicinal chemistry community as a convenient reference that is based upon relevant data. As drug discovery is an iterative process where chemists often consider the impact of a marginal fragmental addition to their lead compound, easily accessible ΔlogD data for the most common substituents can allow for rapid assessment of new targets as opposed to de novo prediction of the lipophilicity of an entire molecule. In addition to lipophilicity, huge reservoirs of data are prevalent in large organizations which can provide value through the application of MMP analysis. Initiatives such as the sharing of data and trends from MMP analyses offer the advantage of being able to inform and enable the academic, pharma, and biotech sectors in a precompetitive manner,^{26,27} and we anticipate the completion and report of similar analyses in the near future.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsmchemlett.9b00489>.

Characterization of compounds 1–6 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: crawford.james@gene.com.

ORCID

James J. Crawford: 0000-0002-6408-8246

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Hao Zheng, Timothy Heffron, Fabio Broccatelli, Michael Siu, Matthew Volgraf, and Paul Beroza for helpful discussions; Wendy Lee, Kewei Xu, and Yanzhou Liu for characterization of compounds 1–6; Eric Low and Newton

Wu for performing analytical experiments; and to Genentech, Roche, and research partners thereof for generating the compounds and gathering the data upon which this analysis is built.

■ REFERENCES

- (1) Waring, M. J. Lipophilicity in drug discovery. *Expert Opin. Drug Discovery* **2010**, *5*, 235–248.
- (2) 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 Delivery Rev.* **1997**, *23*, 3–25.
- (3) Wenlock, M. C.; Austin, R. P.; Barton, P.; Davis, A. M.; Leeson, P. D. A Comparison of Physicochemical Property Profiles of Development and Marketed Oral Drugs. *J. Med. Chem.* **2003**, *46*, 1250–1256.
- (4) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat. Rev. Drug Discovery* **2015**, *14*, 475–486.
- (5) Lin, B.; Pease, J. H. A Novel Method for High Throughput Lipophilicity Determination by Microscale Shake Flask and Liquid Chromatography Tandem Mass Spectrometry. *Comb. Chem. High Throughput Screening* **2013**, *16*, 817–825.
- (6) Fujita, T.; Iwasa, J.; Hansch, C. A New Substituent Constant, π , Derived from Partition Coefficients. *J. Am. Chem. Soc.* **1964**, *86*, 5175–5180.
- (7) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley-Interscience: New York, 1979.
- (8) Hansch, C.; Leo, A. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; ACS: WA, 1995.
- (9) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR. Hydrophobic, Electronic and Steric Constants*; ACS: WA, 1995.
- (10) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525–616.
- (11) Rekker, R. F.; de Kort, H. M. The hydrophobic fragmental constant; an extension to a 1000 data point set. *Eur. J. Med. Chem. Chimica Therapeutica.* **1979**, *14*, 479–488.

- (12) Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity—The Hydrophobic Fragmental Constant Approach*; Wiley-VCH Verlag GmbH: Weinheim, 1992.
- (13) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of Log P Methods on More Than 96,000 Compounds. *J. Pharm. Sci.* **2009**, *98*, 861–893.
- (14) Arnott, J. A.; Planey, S. L. The influence of lipophilicity in drug discovery and design. *Expert Opin. Drug Discovery* **2012**, *7*, 863–875.
- (15) Lo, Y.-C.; Rensi, S. E.; Torng, W.; Altman, R. B. Machine learning in chemoinformatics and drug discovery. *Drug Discovery Today* **2018**, *23*, 1538–1546.
- (16) Dalke, A.; Hert, J.; Kramer, C. mmpdb: An Open-Source Matched Molecular Pair Platform for Large Multiproperty Data Sets. *J. Chem. Inf. Model.* **2018**, *58*, 902–910.
- (17) Gleeson, P.; Bravi, G.; Modi, S.; Lowe, D. ADMET rules of thumb II: A comparison of the effects of common substituents on a range of ADMET parameters. *Bioorg. Med. Chem.* **2009**, *17*, 5906–5919.
- (18) Papadatos, G.; Alkarouri, M.; Gillet, V. J.; Willett, P. Lead Optimization Using Matched Molecular Pairs: Inclusion of Contextual Information for Enhanced Prediction of hERG Inhibition, Solubility, and Lipophilicity. *J. Chem. Inf. Model.* **2010**, *50*, 1872–1886.
- (19) Griffen, E.; Leach, A. G.; Robb, G. R.; Warner, D. J. Matched Molecular Pairs as a Medicinal Chemistry Tool. *J. Med. Chem.* **2011**, *54*, 7739–7750.
- (20) For an MMP analysis of Pfizer lipophilicity data, see: Keefer, C. E.; Chang, G.; Kauffman, G. W. Extraction of tacit knowledge from large ADME data sets via pairwise analysis. *Bioorg. Med. Chem.* **2011**, *19*, 3739–3749.
- (21) Dossetter, A. G.; Griffen, E. J.; Leach, A. G. Matched Molecular Pair Analysis in drug discovery. *Drug Discovery Today* **2013**, *18*, 724–731.
- (22) Tyrchan, C.; Evertsson, E. Matched Molecular Pair Analysis in Short: Algorithms, Applications and Limitations. *Comput. Struct. Biotechnol. J.* **2017**, *15*, 86–90.
- (23) Patani, G. A.; LaVoie, E. J. Bioisosterism: A Rational Approach in Drug Design. *Chem. Rev.* **1996**, *96*, 3147–3176.
- (24) Meanwell, N. A. Synopsis of Some Recent Tactical Applications of Bioisosteres in Drug Design. *J. Med. Chem.* **2011**, *54*, 2529–2591.
- (25) Meanwell, N. A. Improving Drug Design: An Update on Recent Applications of Efficiency Metrics, Strategies for Replacing Problematic Elements, and Compounds in Nontraditional Drug Space. *Chem. Res. Toxicol.* **2016**, *29*, 564–616.
- (26) Kramer, C.; Ting, A.; Zheng, H.; Hert, J.; Schindler, T.; Stahl, M.; Robb, G.; Crawford, J. J.; Blaney, J.; Montague, S.; Leach, A. G.; Dossetter, A. G.; Griffen, E. J. Learning Medicinal Chemistry Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET) Rules from Cross-Company Matched Molecular Pairs Analysis (MMPA). *J. Med. Chem.* **2018**, *61*, 3277–3292.
- (27) Johnstone, C. Medicinal chemistry matters - a call for discipline in our discipline. *Drug Discovery Today* **2012**, *17*, 538–543.