Analysis of Benzenoid Substitution Patterns in Small Molecule Active Pharmaceutical Ingredients

Aleksandra Nilova and David R. Stuart*

Department of Chemistry, Portland State University, Portland, Oregon 97201, United States

dstuart@pdx.edu

Abstract: An analysis of benzenoid substitution patterns in small molecule active pharmaceutical ingredients (APIs) approved by the FDA reveals a preference for 1,4-substituted (*para*), 1-substituted (*mono*), 1,2,4-substituted, and 1,2-substituted (*ortho*) arenes. Notably, these substitution patterns are widely commercially available and readily accessible by electrophilic aromatic substitution (S_EAr), but more highly substituted and contra-electronic substitution patterns are severely underrepresented in drug substances. Finally, structural variation decreases with increasing substitution and there is a strong reliance on natural product scaffolds in drugs with more highly substituted benzenoid rings.

Introduction

Analysis of active pharmaceutical ingredients (API) to highlight features of "druglikeness" and identify priority areas continues to invigorate strategies for discovery and development of small molecule drugs. Previous analyses have focused on physicochemical properties[1] and prominent scaffolds[2] in drugs, and the reagents[3] and reactions[4] used to synthesize them. Broadly, rings have emerged as a scaffold of immense importance for druglikeness,[1a,b,2b,c,d,f,i] and benzenoid rings are, by an order of magnitude, the most frequently encountered ring system in small molecule drugs.[2f] Our interest in the reactivity and synthesis of benzenoid systems and the overwhelming frequency that they occur in APIs prompted us to consider the distribution of benzenoid substitution patterns in small molecule pharmaceuticals. Although a previous analysis by Brown and co-workers found a connection between bias in screening libraries and approved drugs for *para*-chlorophenyl relative to the *meta*- and *ortho*-regioisomers,[3] a complete analysis of benzenoid substitution patterns in approved API is lacking from the literature.

Results and Discussion

We conducted an analysis of all small molecule drugs approved by the FDA up to 2018 using the database Drugbank.ca[5] and cross-referencing to the FDA Orange Book.[6,7] We used the filters "small molecule drugs" for type of drug, "approved" for group, and "US" for market. By visual analysis, we identified 1323 benzenoid rings from the 1793 API "hits" based on the filters. Our analysis excluded benzenoid rings fused to other (hetero)aromatic systems (i.e., naphthalene, (iso)quinoline, indole) because previous analyses identify these as distinct ring systems.[2f] Moreover, such rings are often directly functionalized and therefore are not relevant to a discussion of synthetic strategies to functionalized benzenoid rings. We found that approximately half the APIs (879; 49%) contained benzenoid rings, and of those benzenoid-

containing APIs 60% (524) contained one benzenoid ring, 34% (300) contained two benzenoid rings, 5% (42) contained three benzenoid rings, and 1% (13) contained more than three benzenoid rings (Figure 1).

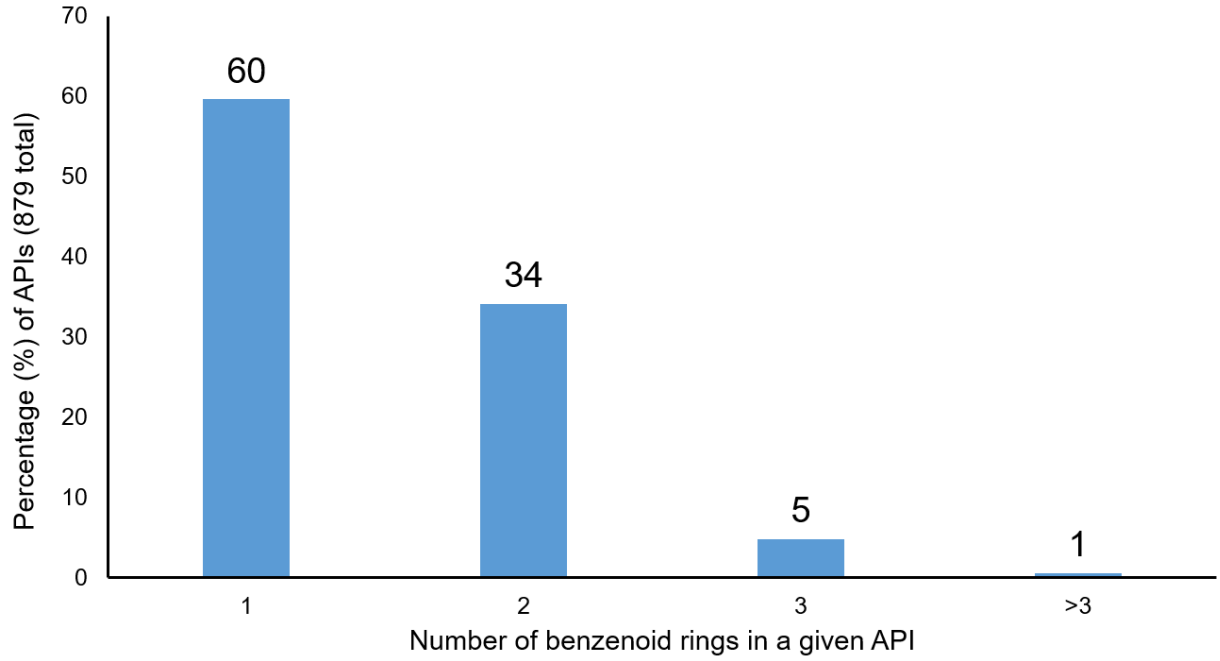


Figure 1. Distribution of the number of benzenoid rings in a given API

Of the 1323 benzenoid rings analyzed, the percentage of each substitution pattern is shown in Figure 2. In our analysis we are considering a total of twelve different substitution patterns irrespective of the substituents identity. By far, the most prevalent benzenoid substitution patterns are 1,4-substituted (23%), 1-substituted (23%), 1,2,4-substituted (18%), and 1,2-substituted (12%). These four substitution patterns constitute 76% of all benzenoid ring substitution patterns, and are those most easily accessed by S_EAr (aromatic halogenation) and are the most widely commercially available aromatic starting materials.[4a] The remaining 24% of benzenoid substitution patterns found in FDA approved small molecule drugs are spread across the other eight different substitution patterns and the individual percentages range from 1-5%. Notably, the arrangement of substituents in the minority of substitution patterns are *contra*-electronic and deactivated to S_EAr (i.e., 1,3- and 1,3,5-) or are sterically congested (i.e., 1,2,3-; 1,2,3,5-; 1,2,3,4-; etc). Building blocks with these patterns are less commercially available and a greater synthetic investment is required to access them. Previous analyses have noted their absence from screening libraries and commercial databases,[3,4a] and Carey *et al*. explicitly stated *"These are seemingly simple starting materials that are often very difficult to produce. New methods for the synthesis of these difficult substitution patterns would be welcomed."*.[4a]

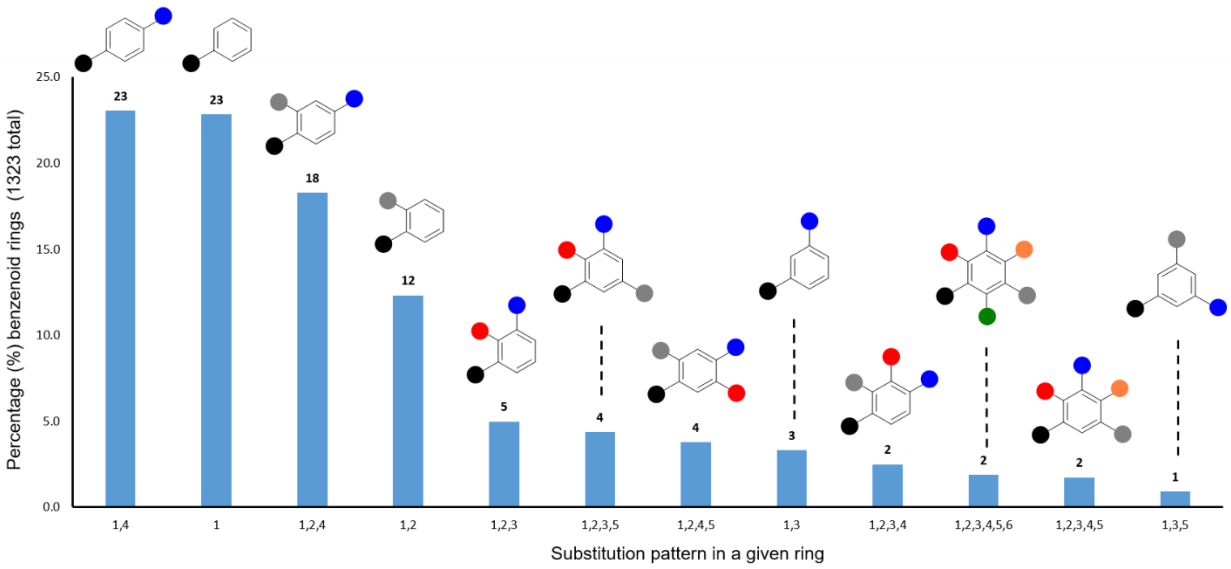


Figure 2. Distribution of benzenoid substitution patterns in APIs.

We looked deeper for a possible correlation between benzenoid ring substitution patterns and time given the ever increasing advent of new synthetic methods. A previous analysis by Brown and Boström investigated which reactions are most commonly used in medicinal chemistry,[4d] and those associated with reactions at benzenoid carbons were S_NAr, Suzuki-Miyaura coupling, aryl lithiation and electrophilic trapping, Sonogashira coupling, aromatic halogenation (S_EAr), and Buchwald-Hartwig amination. Brown and Boström also compared reactions used by medicinal chemists in 1984 and 2014 and the clear difference between these two data sets was the appearance of palladium-catalyzed coupling reactions being used much of the time in 2014. Suzuki-Miyaura coupling[8] was the 5th most commonly used reaction overall, and there were two other palladium-catalyzed reactions (Sonogashira[9] and Buchwald-Hartwig[10]) in the top twenty. Indeed, palladium-catalyzed reactions have had a dramatic impact on synthetic chemistry,[11] including medicinal chemistry. In our analysis we compared two data sets: 1) substitution patterns in drugs disclosed from 1958-88, and 2) substitution patterns in drugs disclosed from 1998-2018 (Figure 3).[12] Consistent with the study of Brown and Boström, the introduction of palladium catalysis increased the number of benzenoid rings in drugs from 35 in 1958-1988 to 90 in 1998-2018. However, Figure 3 shows that benzenoid substitution patterns did not substantially change between these two time periods and are consistent with the overall distribution in Figure 2. Broadly, the most common substitution patterns in both time periods were 1,4-; 1-; 1,2,4-; and 1,2-substitution rings, constituting 75% in 1958-1988 and 62% in 1998-2018.

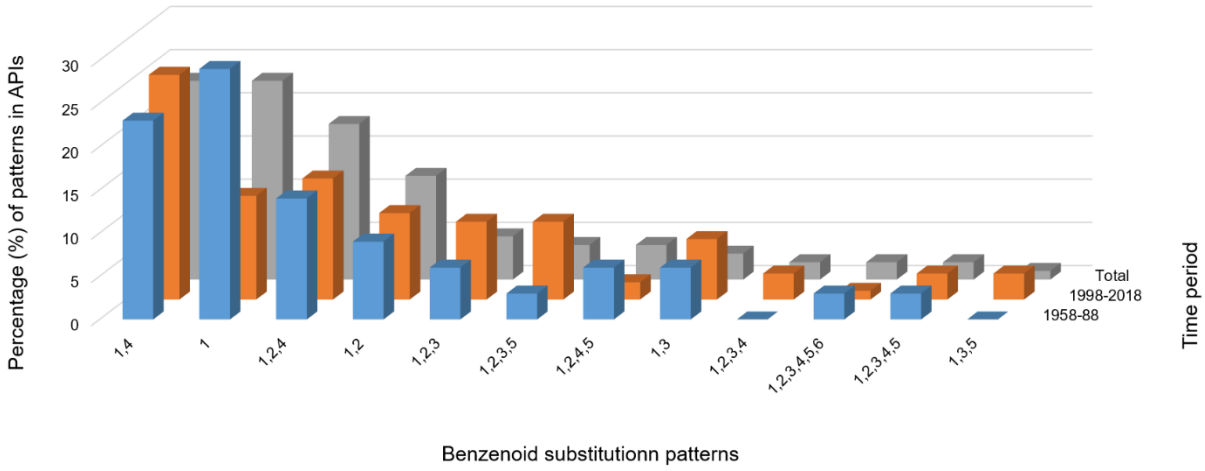


Figure 3. Distribution of benzenoid substitution patterns across several time periods.

A structural analysis of underrepresented benzenoid patterns is informative and underscores synthetic challenges. Figure 4 shows selected examples of the eight underrepresented benzenoid substitution patterns. Overall, there appears to be an inverse relationship between structural variation and number of substituents. Although a greater number of substituents on a benzenoid rings should lead to greater structural variation, the synthetic "cost" to install more substituents has the opposite effect. Consequently, there also appears to be a greater reliance on natural product scaffolds for more highly and densely substituted benzenoid rings. For instance, we analyzed a total of 33 benzenoid rings across 31 unique APIs with the 1,2,3,4-substituted benzenoid ring family; of those ~50% (15) are morphine-derived compounds wherein the 1,2,3,4-substitution pattern on the A-ring is installed by Mother Nature. Additionally, within the 1,2,3,5-substituted rings (58 total), the gallic acid moiety (i.e., 3,4,5-trihydroxybenzoic acid, or a variant thereof) features prominently (10 examples). Finally, there is also relatively little structural variation for the 1,2,3,4,5,6-substituted benzenoids wherein ~ 75% are structurally related to iothalamic acid, alpha-tocopherol (Vitamin E), or doxorubicin, with the latter two being natural products.

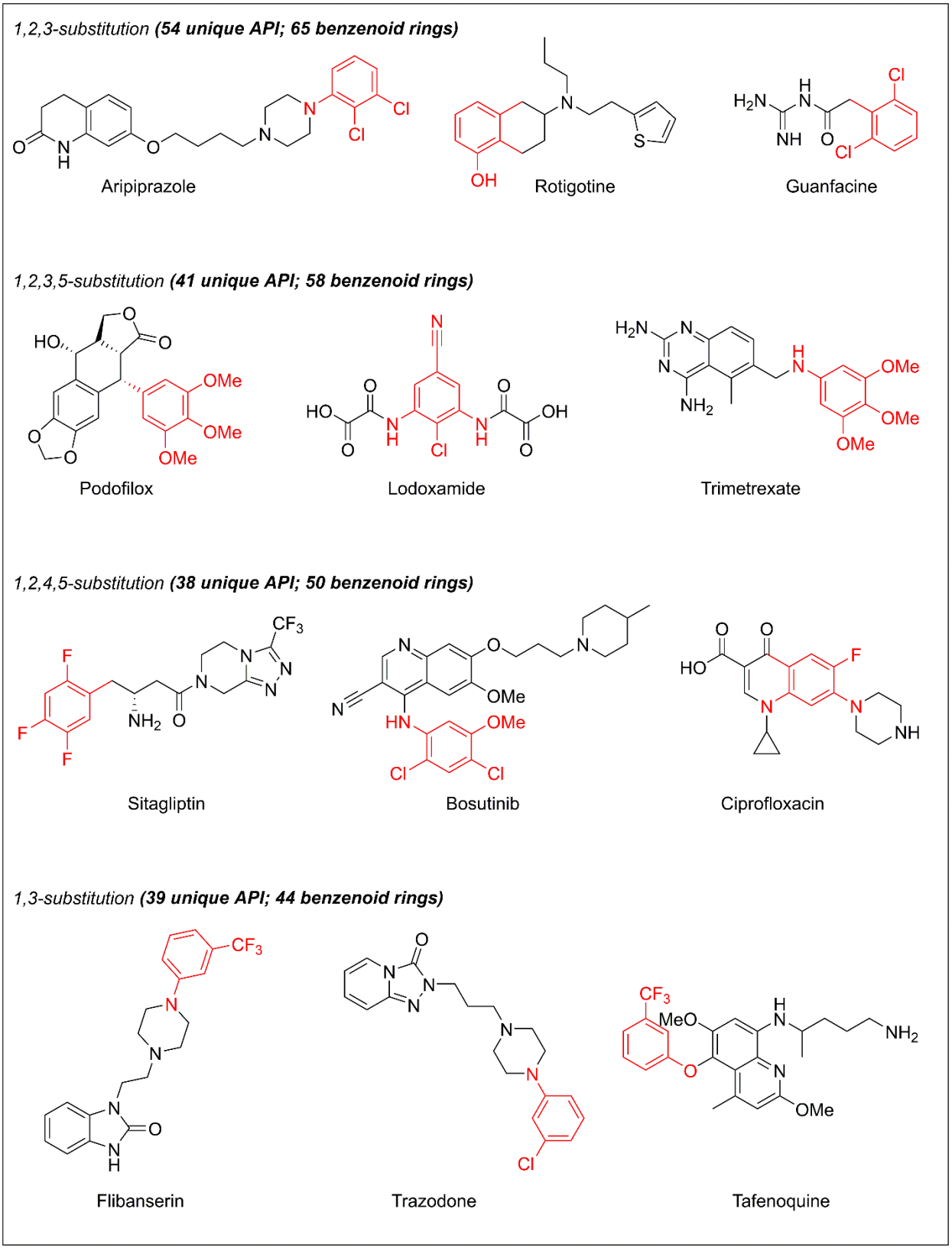
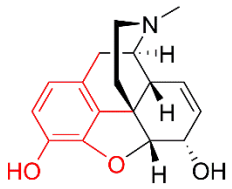
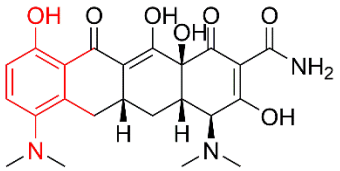


Figure 4. Representative examples of underrepresented benzenoid substitution patterns.

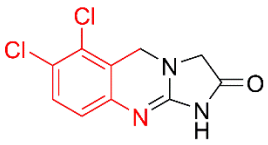
1,2,3,4-substitution ***(31 unique API; 33 benzenoid rings)***



Morphine

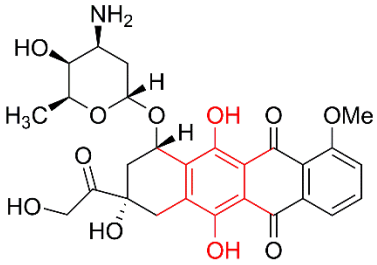


Minocycline

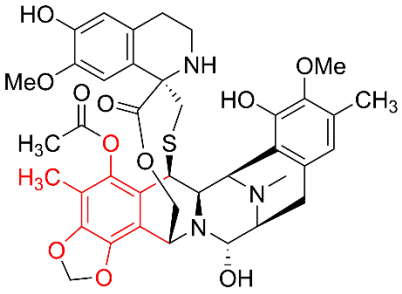


Anagrelide

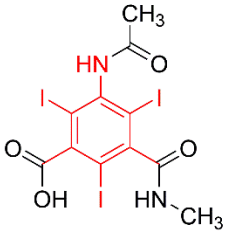
1,2,3,4,5,6-substitution ***(23 unique API; 25 benzenoid rings)***



Doxorubicin

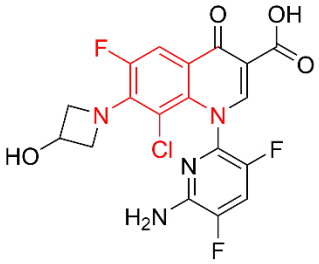


Trabectedin

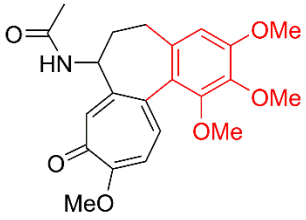


Iothalamic acid

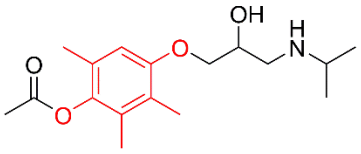
1,2,3,4,5-substitution ***(21 unique API; 23 benzenoid rings)***



Delafloxacin

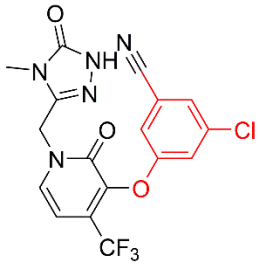


Colchicine

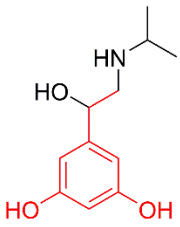


Metipranolol

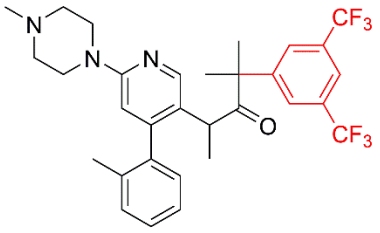
1,3,5-substitution ***(12 unique API; 12 benzenoid rings)***



Doravirine



Orciprenaline



Netupitant

Figure 4 cont.

Conclusion

We have analyzed all small molecule APIs approved by the FDA and tabulated the substitution patterns found in benzenoid rings. We observed a strong preference (76%) for S_EAr-type patterns (1-; 1,4-; 1,2-; and 1,2,4-substitution), and our findings are consistent with previous analyses that have focused on reagents and reactions used in medicinal chemistry.[3,4] In our opinion the bias in benzenoid substitution patterns is, in part, attributed to the fact that synthetic methods to access contra-electronic or sterically congested benzenoid rings have not kept pace with methods that replace substituents on S_EAr patterns, and this bias will continue unless broken by the invention of new synthetic methods that address this challenge.

Acknowledgments

We thank Portland State University for financial support.

References

[1] (a) Gibson, S.; McGuire, R.; Rees, D. C. Principal Components Describing Biological Activities and Molecular Diversity of Heterocyclic Aromatic Ring Fragments. *J. Med. Chem.* **1996**, *39*, 4065-4072. (b) Ritchie, T. J.; Macdonald, S. J. F. The Impact of Aromatic Ring Count on Compound Developability – Are too Many Aromatic Rings a Liability in Drug Design? *Drug Discovery Today* **2009**, *14*, 1011-1020. (c) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752-6756.

[2] (a) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887-2893. (b) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; Mclay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface. A Tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257-3275. (c) Ertl, P.; Jelfs, S.; Mühlbacher, J.; Schuffenhauer, A.; Selzer, P. Quest for the Rings. In Silico Exploration of Ring Universe to Identify Novel Bioactive Heteroaromatic Scaffolds. *J. Med. Chem.* **2006**, *49*, 4568-4573. (d) Pitt, W. R.; Parry, D. M.; Perry, B. G.; Groom, C. R. Heteroaromatic Rings of the Future. *J. Med. Chem.* **2009**, *52*, 2952-2963. (e) Wang, J.; Hou, T. Drug and Drug Candidate Building Block Ananlysis. *J. Chem. Inf. Model.* **2010**, *50*, 55-67. (f) Taylor, R. D.; MacCross, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845-5859. (g) Ilardi, E. A.; Vitaku, E.; Njardarson, J. T.; Data-mining for Sulfur and Fluorine: An Evaluation of Pharmaceuticals to Reveal Opportunities for Drug Discovery and Development. *J. Med. Chem*. **2014**, *57*, 2832. (h) Smith, B. R.; Eastman, C. M.; Njardarson, J. T. Beyond C, H, O, and N! Analysis of the Elemental Composition of US FDA Approved Drug Architechtures. *J. Med. Chem.* **2014**, *57*, 9764. (i) Vitaku, E.; Smith, D. T.; Njardarson, J. T. Analysis of the Structural Diversity, Substitution Patterns and Frequency of Nitrogen Heterocycles among US FDA Approved Pharmaceuticals. *J. Med. Chem.* **2014**, *57*, 10257. (j) Delost, M. D.; Smith, D. T.; Anderson, B. J.; Njardarson, J. T. From Oxiranes to Oligomers: Architectures of US FDA Approved Pharmaceuticals Containing Oxygen Heterocycles. *J. Med. Chem.* **2018**, *61*, 10996. (k) Das, P.; Delost, M.; Qureshi, M.; Smith, D. T.; Njardarson, J. T. A Survey of the Structures of US FDA Approved Combination Drugs. *J. Med. Chem.* 10.1021/acs.jmedchem.8b01610

[3] Brown, D. G.; Gagnon, M. M.; Boström, J. Understanding Our Love Affair with *p*-Chlorophenyl: Present Day Implications from Historical Biases of Reagent Selection. *J. Med. Chem.* **2015**, *58*, 2390-2405.

[4] (a) Carey, J. S.; Laffan, D.; Thomson, C.; Williams, M. T. Analysis of the Reactions Used for the Preparation of Drug Candidate Molecules. *Org. Biomol. Chem.* **2006**, *4*, 2337-2347. (b) Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *J. Med. Chem.* **2011**, *54*, 3451-3479. (c) Walters, W. P.; Green, J.; Weiss, J. R.; Murcko, M. A. What Do Medicinal Chemists Actually Make? A 50-Year Retrospective. *J. Med. Chem.* **2011**, *54*, 6405-6416. (d) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2015**, *59*, 4443-4458.

[5] Searched www.drugbank.ca/drugs on 01/31/2019.

[6] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2017 Nov 8. doi: 10.1093/nar/gkx1037.

[7] https://www.accessdata.fda.gov/scripts/cder/ob/index.cfm

[8] Miyaura, N.; Yanagi, T.; Suzuki, A. The Palladium-Catalyzed Cross-Coupling Reaction of Phenylboronic Acid with Haloarenes in the Presence of Bases. *Synth. Commun.* **1981**, *11*, 513-519.

[9] Sonogashira, K.; Tohda, Y.; Hagihara, N. A Convenient Synthesis of Acetylenes: Catalytic Substitutions of Acetylenic Hydrogen with Bromoalkenes, Iodoarenes, and Bromopyridines. *Tetrahedron Lett.* **1975**, *16*, 4467-4470.

[10] (a) Paul, F.; Patt, J.; Hartwig, J. F. Palladium-Catalyzed Formation of Carbon-Nitrogen Bonds. Reaction Intermediates and Catalyst Improvements in the Hetero Cross-Coupling of Aryl Halides and Tin Amides. *J. Am. Chem. Soc.* **1994**, *116*, 5969-5970. (b) Guram, A. S.; Buchwald, S. L. Palladium-Catalyzed Aromatic Aminations with in situ Generated Aminostannanes. *J. Am. Chem. Soc.* **1994**, *116*, 7901-7902

[11] (a) Suzuki, A. Cross-Coupling Reactions of Organoboranes: An Easy Way to Construct C-C Bonds (Nobel Lecture) *Angew. Chem. Int. Ed.* **2011**, *50*, 6723-6737. (b) Negishi, E. Magical Power of Transition Metals: Past, Present, and Future (Nobel Lecture) *Angew. Chem. Int. Ed.* **2011**, *50*, 6738-6764.

[12] In one set we combined the substitution patterns observed in the years 1958, 1968, 1978, and 1988 (35 total rings); and in the other data set we combined the substitution patterns observed in the years 1998, 2008, and 2018 (90 total rings).