molecules

MDPI

Article

How Size Matters: Diversity for Fragment Library Design

Yun Shi * and Mark von Itzstein *

Institute for Glycomics, Griffith University, Gold Coast Campus, Gold Coast, Queensland 4222, Australia
* Correspondence: y.shi@griffith.edu.au (Y.S.); m.vonitzstein@griffith.edu.au (M.v.I.)

Academic Editors: Diego Muñoz-Torrero, F. Javier Luque and Marçal Pastor-Anglada
Received: 18 July 2019; Accepted: 3 August 2019; Published: 5 August 2019

Abstract: Fragment-based drug discovery (FBDD) has become a major strategy to derive novel lead candidates for various therapeutic targets, as it promises efficient exploration of chemical space by employing fragment-sized (MW < 300) compounds. One of the first challenges in implementing a FBDD approach is the design of a fragment library, and more specifically, the choice of its size and individual members. A diverse set of fragments is required to maximize the chances of discovering novel hit compounds. However, the exact diversity of a certain collection of fragments remains underdefined, which hinders direct comparisons among different selections of fragments. Based on structural fingerprints, we herein introduced quantitative metrics for the structural diversity of fragment libraries. Structures of commercially available fragments were retrieved from the ZINC database, from which libraries with sizes ranging from 100 to 100,000 compounds were selected. The selected libraries were evaluated and compared quantitatively, resulting in interesting size-diversity relationships. Our results demonstrated that while library size does matter for its diversity, there exists an optimal size for structural diversity. It is also suggested that such quantitative measures can guide the design of diverse fragment libraries under different circumstances.

Keywords: diversity; fragment-based drug discovery; library design; library size

1. Introduction

Fragment-based drug discovery (FBDD) has been developed in the past twenty years as an approach to derive novel lead compounds for various therapeutic targets [1–4]. It features the use of fragment-sized compounds that mostly comply with the 'Rule-of-3' [5] for the identification of hits, which can be subsequently developed into potent lead compounds. Compared to the more traditional high-throughput screening that employs drug-like compounds following the 'Rule-of-5' [6], the smaller sizes of fragments used in FBDD lead to more efficient sampling of the relevant chemical space and thus better chances of identifying novel hits [7]. The smaller sizes also result in higher ligand efficiency [8] and more efficient structural optimization of fragment hits [9]. With these advantages, FBDD has gained popularity in both academia and industry in recent years [10], and led to the discovery of three Food and Drug Administration-approved drugs [11–14].

The first and foremost step in FBDD is the design of a fragment library, as library compositions directly influence the outcome of FBDD projects. One of the most frequently discussed topics for its design is the size of the fragment library, which has a substantial impact on the early stages as it affects the time and monetary costs in addition to the outcome of FBDD projects. Interestingly, the majority of respondents in recent polls had up to 2,000 compounds in their fragment libraries [15,16], while recent successful FBDD campaigns had library sizes of between 1,000 and 2,000 compounds [17,18]. Besides an optimal library size, consideration is also given to the structural complexity, physicochemical profile, and shape profile of fragments [19]. However, the *diversity* of a fragment library should be the most

critical factor, because it affects the sampling efficiency of the relevant chemical space as well as the novelty of potential hit compounds. Of note, better diversity should decrease the screening hit rate, which appears to be high for many FBDD campaigns [4,20]. Hence, the size of a fragment library should be discussed in conjunction with its diversity.

Diversity needs to be characterized by descriptors, which can be classified mostly into three categories. The first are functional (performance) descriptors based on the bioactivities of compounds towards a panel of (functionally dissimilar) biological targets [21]. Although regarded as the most relevant category of diversity descriptors for drug discovery [22,23], acquisition of bioactivity data can be very resource-demanding [24,25]. In addition to a lack of bioactivity data for fragment-sized compounds in the literature, their activities would also be difficult to detect and measure due to their weak affinities [7]. The second are physicochemical (property-based) descriptors, including common physicochemical properties such as molecular weight, hydrophobicity, and even electronic properties [26]. The third are structural descriptors, among which molecular fingerprints (structural features) are routinely used to represent chemical structures. The extended-connectivity (radial) fingerprints [27] is effective at retrieving bioactive compounds [28], therefore it was chosen as the descriptor of diversity in our study.

There are currently two major types of quantitative metrics for structural diversity [29]. The first type of metrics assesses the similarity (and thus difference) between pairs of chemical structures. The most notable metric of this type is the Jaccard index [30], later referred to as the popular Tanimoto index (similarity) [31]. The second type of metrics calculates the coverage of the relevant chemical space by a library of compounds, and the most straightforward one is a ratio based on richness, defined as the number of unique fingerprints (structural features) [32]. In this work, we propose the adoption of a third type of metrics, i.e., a diversity index that takes into account not only the number of unique structural fingerprints but also their proportional abundances [33–35], for the quantitative measurement of diversity. True diversity, or the effective number of structural features, is a commonly used metric of this type and can be defined by the following Equation (1) [35]:

$$D = \frac{1}{\prod_{i=1}^{R} p_i^{p_i}} \quad (1)$$

where D stands for true diversity, R is richness (the total number of fingerprints), and p_i represents the proportional abundance of the ith fingerprint. It can be deduced from Equation (1) that, for the same richness, a library with a more even distribution of proportional abundances will have a larger true diversity than a library with a less even distribution. These diversity indexes have been used in ecological studies for decades, yet they have not been applied to the measurement of diversity of fragment libraries to date. Although there are other plot-based methods to illustrate diversity in more visually appealing ways, such as principle component analysis [36] and principal moments of inertia [37,38], these three quantitative metrics, i.e., Tanimoto similarity, number of fingerprints, and true diversity, are more suited for direct comparison of libraries with different sizes.

To provide insights into how the library size affects the structural diversity, we herein compare fragment libraries of different sizes, selected from commercially available fragments, and demonstrate interesting size-diversity relationships. Such relationships indicated the presence of an optimal library size for structural diversity. We also extend this investigation to a more restrictive scenario, in which only fluorinated fragments are considered and consequentially similar size-diversity relationships were observed. Certain cost-effective sizes that capture significant proportions of the overall diversity available with very small portions of available fragments are also proposed. Our results demonstrated that these quantitative metrics could assist in the design of fragment libraries under various circumstances.

2. Results

2.1. Library Selection

To generate libraries for comparison, both diversity-based selections and random selections were performed from 227,787 commercially available fragments that had undergone filtering by the 'Rule-of-3' criteria [5]. Libraries with sizes of 100, 200, 500, 1,000, 2,000, 5,000, 10,000, 20,000, 50,000, and 100,000 were selected. Both diversity-based selections and random selections were performed, with the latter in triplicate. To demonstrate that our approach can be applied to different circumstances, selections were also performed on a fluorinated subset of the 227,787 commercially available fragments, consisting of 47,708 fragments that has 1~3 fluorine atoms. Such restriction on the number of fluorine atoms captured the majority of fluorinated fragments, which are commonly used for FBDD projects employing ^{19}F NMR as the screening method [39,40]. Fluorinated libraries with sizes of 100, 200, 500, 1,000, 2,000, 5,000, 10,000, and 20,000 were selected in similar fashions.

2.2. Size-Diversity Relationship of Regular Fragment Libraries

To understand the relationship between the size of fragment libraries and their structural diversity, quantitative metrics were calculated for selected libraries (Figure 1). As expected, fragments became more similar to each other as the library size increased, and the diversity-based selection did lead to more dissimilar fragments than random selections (Figure 1a). Richness of fragment library also rose with its size, with diversity-based selections outperforming random selections (Figure 1b). However, marginal richness, i.e., the additional number of unique fingerprints per additional fragment, was declining while library size grew (Figure 2a). For diversity-based selections, the average efficiency of adding unique fingerprints from 2,000 fragments to 5,000 fragments, 13.4 fingerprints per compound, was less than half of that from nothing to 100 fragments, 28.9 fingerprints per compound. Similar trends were observed for randomly selected libraries, although the gap between diversity-based and random selections became smaller when library sizes grew excessively large, i.e., beyond 5,000 compounds. Thus, it is more efficient to have relatively small library for richness and we estimated the number of fragments required to accomplish two arbitrary degrees of coverage, 5% and 10%, respectively (Table 1). These two cut-offs are convenient numbers chosen to manifest the coverage efficiency of small libraries.

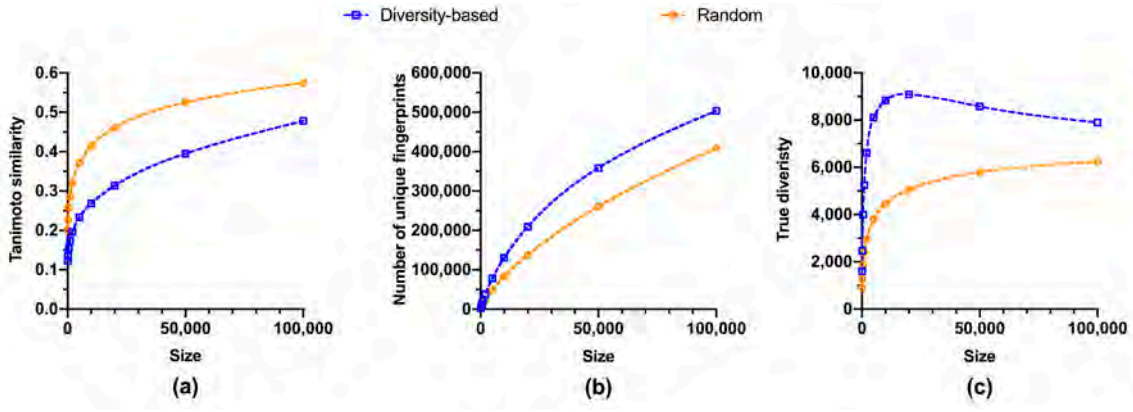

Figure 1. Structural diversity vs size of fragment libraries, with the former measured by: (**a**) Average of the similarity of each compound to its closest neighbor; (**b**) total number of unique fingerprints (richness); (**c**) true diversity calculated by equation 1. Dash curves are generated from cubic spline fitting. Metrics for random selections are average values of triplicates (Table S2).

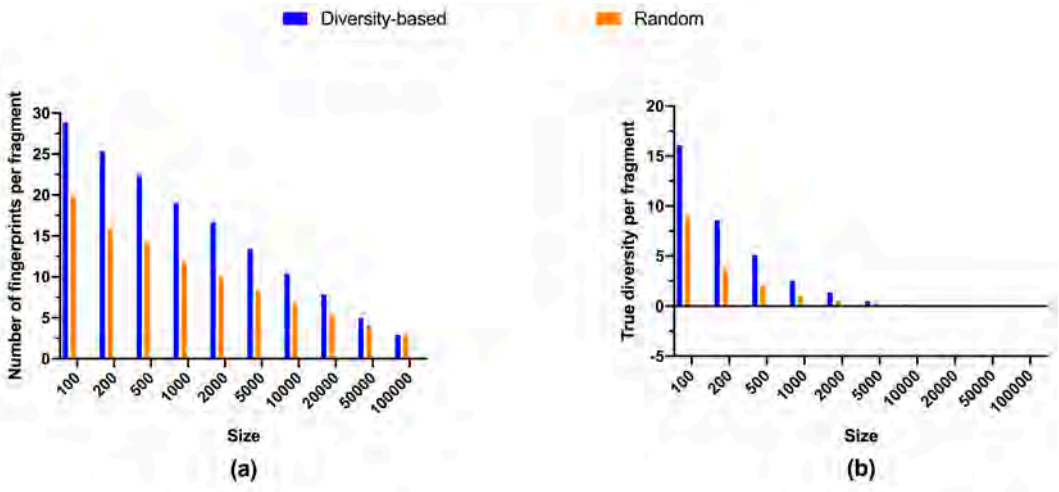

Figure 2. Efficiency in adding diversity: (**a**) average number of unique fingerprints (richness) per compound; (**b**) average value of true diversity per compound. Metrics for random selections are average values of triplicates.

Table 1. Library sizes (diversity-based selection) required to achieve certain values of structural diversity.

Structural Diversity (Value)	Minimum Size (Ratio of Total 227,787 Fragments) [1]
5% total richness [2] (33,834)	1,715 (0.75%)
10% total richness [2] (67,669)	4,103 (1.80%)
Overall true diversity (6,662.4)	2,052 (0.90%)
Maximum true diversity [1] (9,097.6)	17,666 (7.76%)

[1] Values are estimated by cubic spline fitting with 99,901 segments; [2] Total richness (number of unique fingerprints) is 676,686.

Surprisingly, values of true diversity exhibited different trends between diversity-based selections and random selections (Figure 1c). While the latter showed a constantly rising movement, the former reached a maximum at about 18,000 fragments, representing less than 8% of the overall available fragments (Table 1), before starting to decline (Figure 1c). In addition, marginal true diversity experienced a more drastic decline in comparison with the marginal richness (Figure 2). For diversity-based selections, the average efficiency of adding true diversity from 2,000 fragments to 5,000 fragments, 1.4 per compound, was an order of magnitude less that from nothing to 100 fragments, 16.1 per compound. Consistent with the decline of true diversity after the library size from diversity-based selections reached about 18,000, the marginal true diversity became negative after 20,000 compounds (Figure 2b). More strikingly, only approximately 2,000 fragments, i.e., less than 1%, are required to attain the same level of true diversity as all of the 227,787 fragments available for selection (Table 1).

2.3. Size-Diversity Relationship of Fluorinated Fragment Libraries

Libraries selected from fluorinated fragments presented similar size-diversity relationships as those from regular fragments (Figures 3 and 4, Table 2). Both similarity to the closest neighbor and richness illustrated growing trends (Figure 3a,b), whereas the true diversity for libraries subject to diversity-based selection also reached a maximum at about 7,500 fragments (Figure 3c and Table 2). Analogously, both marginal richness and marginal true diversity diminished with increasing library size, while the gap between diversity-based and random selections in efficiency became smaller for larger library sizes, i.e., beyond 500 compounds (Figure 4). Nevertheless, it required relatively more fluorinated fragments to achieve the same level of diversity than that for regular fragments. About 3.4% of total fluorinated fragments were needed to attain 10% coverage (Table 2), much higher than that for regular fragments, about 1.8%. Additionally, it took close to 15.7% of total fluorinated fragments to

reach maximum true diversity, while for regular fragments only about 7.8% were required. Further, 2.5% of total fluorinated fragments were required to achieve the same level of true diversity as all the 47,708 fluorinated fragments, whereas less than 1% of regular fragments were required. These observations can be explained by the constant presence of fluorine atoms, and thus fluorine-containing fingerprints, in all fluorinated compounds. Inevitably, there would be a larger overlap of fluorine-associated fingerprints among fluorinated compounds, rendering the distribution of proportional abundances for fingerprints less even and thereby a smaller value of true diversity calculated by equation 1. Such a phenomenon can also be expected for other restrictive circumstances demanding the presence of certain functional groups and/or pharmacophores.

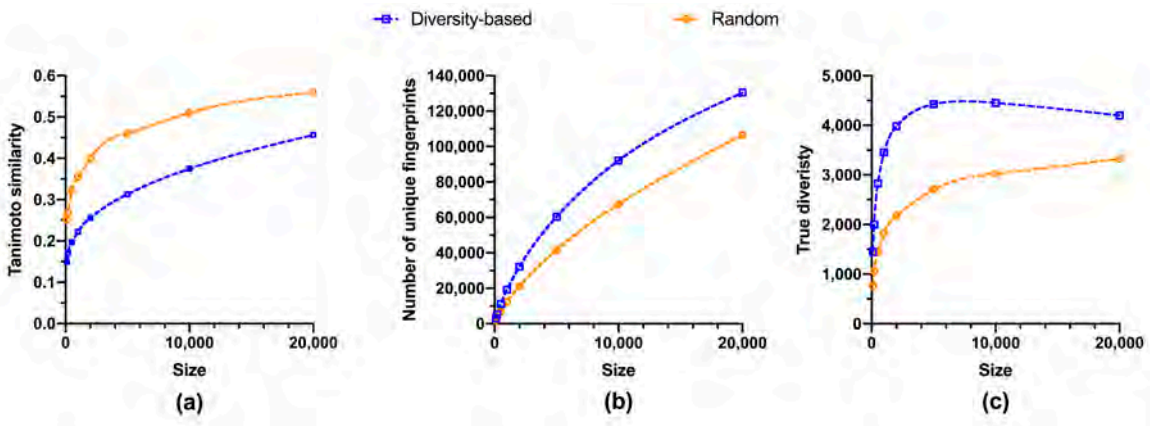

Figure 3. Structural diversity vs size of fluorinated fragment libraries, with the former measured by: (**a**) Average of the similarity of each compound to its closest neighbor; (**b**) total number of unique fingerprints (richness); (**c**) true diversity calculated by equation 1. Dash curves are generated from cubic spline fitting. Metrics for random selections are average values of triplicates (Table S2).

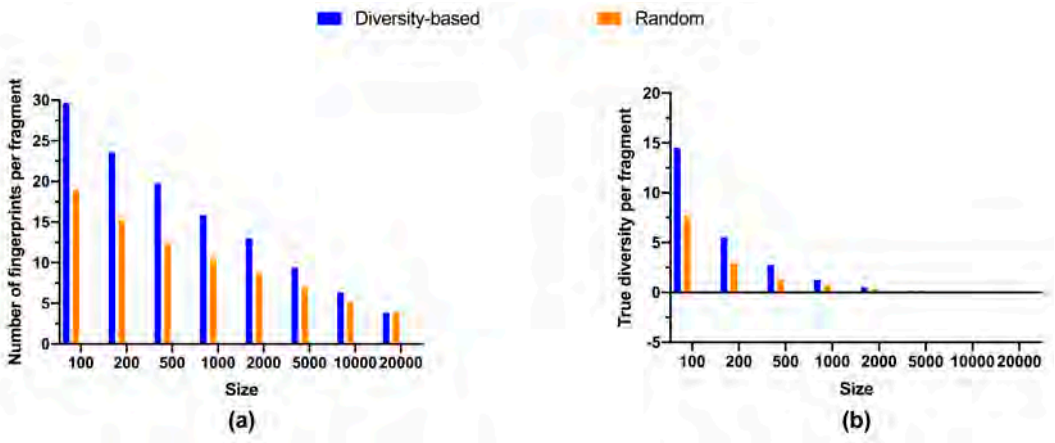

Figure 4. Efficiency in adding diversity: (**a**) average number of unique fingerprints (richness) per fluorinated compound; (**b**) average value of true diversity per fluorinated compound. Metrics for random selections are average values of triplicates.

Table 2. Fluorinated library sizes (diversity-based selection) required to achieve certain values of structural diversity.

Structural Diversity (Value)	Minimum Size (Ratio of Total 47,708 Fluorinated Fragments) [1]
5% total richness [2] (8,992)	675 (1.41%)
10% total richness [2] (17,983)	1,616 (3.39%)
Overall true diversity (3,621.9)	1,203 (2.52%)
Maximum true diversity [1] (4,485.5)	7,483 (15.69%)

[1] Values are estimated by cubic spline fitting with 19,901 segments; [2] Total richness (number of unique fingerprints) is 179,833.

3. Discussion

The exact size-diversity relationships for fragment libraries are affected by several factors, including the fragments available for selection, the selection method, and the diversity metric. Using fluorinated fragments as an example, we have shown that similar size-diversity relationships are observed for this subset of available fragments. Thus, we speculate that different but similar size-diversity relationships could be observed for a different set of fragments available for selection. This could be either more restrictive, such as a set of fragments from a certain vendor, or more inclusive, such as a virtual set of all theoretically possible fragments [41]. Moreover, we expect that a different selection method, such as a clustering method [42], would offer somewhat different results (Table S1). Yet it should be noted that clustering methods are much less efficient than the directed sphere exclusion method used in this study [43], which features good computational performance on large data sets and enabled our calculations to be carried out on a desktop computer. Furthermore, our results illustrated that different diversity metrics could indeed show very different size-diversity relationships. While both similarity and richness increased with the size of fragment library, the rate of increase experienced a more significant decline in the former than in the latter, resulting in larger curvatures of the fitted lines for similarity. In contrast, the true diversity of libraries from diversity-based selections started to decrease after a certain size, highlighting the uneven distribution of structural fingerprints as the library size grew excessively large.

Not unexpectedly, our results showed that the marginal diversity diminishes while the library size increases, the extent and significance of which depends on the choice of diversity metrics. This indicates that it is unnecessary and possibly counterproductive to play numbers game and build excessively large libraries, and that cost-effective sizes of fragment library exist for structural diversity. For regular fragments selected from commercially readily available compounds, we propose a library size of ~2,000 (File S2), corresponding to 0.9% of total available fragments in this study. This size covers more than 5% of richness, approximates the true diversity of all available fragments, and (perhaps coincidentally) matches the most popular fragment library size [15,16]. For the fluorinated subset, a library size of ~1,200 (File S3) achieves similar coverage of richness and true diversity. However, better selection methods may even reduce these proposed numbers.

In addition to structural diversity, considerations should also be given to practical factors such as experimental solubility, (absence of) aggregation, and stability for fragment library design [19]. These factors are essential for the success of FBDD campaigns, yet they are difficult to predict without experimental data. Hence, it would be more pragmatic to slightly increase the library size in the initial *in silico* design and perform necessary quality checks after procurement of fragments.

In summary, we have introduced quantitative metrics to evaluate the structural diversity of fragment libraries, investigated their size-diversity relationships, and demonstrated the existence of an optimal library size for structural diversity depending on specific situations. Based on our results, we propose the use of relatively small library sizes and the application of these quantitative measures to the design of diverse fragment libraries under various circumstances.

4. Materials and Methods

Structures of commercially-available, fragment-sized compounds were retrieved from the ZINC 15 database [44] (https://zinc15.docking.org/tranches/home/) in SMILES format on 2 Jan 2019. A subset was chosen with the following criteria: Anodyne for Reactivity; In-Stock for Purchasability; up to 300 Daltons for Molecular Weight; up to 3 for LogP. These criteria resulted in 1,413,973 compounds. The Canvas program (Schrödinger, LLC, New York, NY, USA) was used for subsequent calculations. Physicochemical properties were calculated by canvasMolDescriptors and compounds violating an adapted version of the 'Rule-of-3' [5], i.e., 100 ≤ MW ≤ 300, logP ≤ 3, number of rings ≤ 3, number of hydrogen bond donors (HBD) ≤ 3, number of hydrogen bond acceptors (HBA) ≤ 3, number of rotatable bonds (RB) ≤ 3, and polar surface area ≤ 60 Å^2 were removed. HBD, HBA, and RB are custom defined according to a previous work [45]. Any compound with reactive groups was filtered by the ligfilter functionality and duplicate structures were eliminated by the uniquesmiles functionality. Finally, 227,787 compounds (File S1 and Figure S1) were left for selection of fragment libraries.

Radial fingerprints [27] were generated by canvasFPGen, with 64-bit precision (2^{64}) to avoid fingerprint collisions, Daylight invariant atom types [46], and three radial iterations. Based on the these fingerprints, diversity-based selections were performed with canvasDBCS, using the directed sphere exclusion method [43] and Tanimoto similarity [31]. An exclusion sphere size of 0.4 was used to select libraries with a maximum size of 100,000 compounds. In parallel, random selections of fragment libraries as control were carried out in triplicate by the UNIX command *shuf*. To quantify the diversity of selected libraries, three different metrics were calculated as follows: maximum Tanimoto similarity [31] was computed by canvasFPHist; total number of unique fingerprints [32] was counted by canvasFPBinary2CSV; and true diversity [35] was determined by the UNIX command *awk* using Equation (1).

For fluorinated fragments, the ligfilter functionality was used to filter the 227,787 compounds with a criterion of 1 ≤ number of fluorine atoms ≤ 3, and the resulting 47,708 fragments were subject to analogous calculations and selections with a maximum library size of 20,000 compounds.

Prism 8 (GraphPad Software, Inc., La Jolla, CA, USA) was employed to generate plots of the aforementioned three metrics against the size of selected libraries, and the cubic spline function was used to fit spine curves.

Supplementary Materials: The following are available online at http://www.mdpi.com/1420-3049/24/15/2838/s1, File S1: List of SMILES structures of all 227,787 fragments used in this study; File S2: List of SMILES structures of 2,000 regular fragments proposed for cost-effectiveness; File S3: List of SMILES structures of 1,200 fluorinated fragments proposed for cost-effectiveness; Table S1: Comparison of clustering-based selections and diversity-base selections from a random set of 10,000 fragments; Table S2: Numerical values of diversity metrics calculated for all selected libraries; Figure S1: Chemical structures of 40 example fragments randomly selected from the 227,787 compounds used for library selections.

Author Contributions: Conceptualization, Y.S. and M.v.I.; methodology, Y.S.; software, Y.S.; formal analysis, Y.S.; investigation, Y.S.; data curation, Y.S.; writing—original draft preparation, Y.S.; writing—review and editing, Y.S. and M.v.I.; visualization, Y.S.; supervision, M.v.I.; funding acquisition, Y.S. and M.v.I.

Funding: This research was funded by the National Health and Medical Research Council, grant number 1071659 (M.v.I.). Y.S. was a recipient of Griffith University Postdoctoral Fellowship.

Acknowledgments: We gratefully acknowledge the support of the Griffith University eResearch Services Team and the use of the High Performance Computing Cluster "Gowonda" to complete this research.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Shuker, S.B.; Hajduk, P.J.; Meadows, R.P.; Fesik, S.W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534. [CrossRef] [PubMed]
2. Erlanson, D.A.; McDowell, R.S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem.* **2004**, *47*, 3463–3482. [CrossRef] [PubMed]
3. Rees, D.C.; Congreve, M.; Murray, C.W.; Carr, R. Fragment-based lead discovery. *Nat. Rev. Drug Discov.* **2004**, *3*, 660–672. [CrossRef] [PubMed]
4. Erlanson, D.A.; Fesik, S.W.; Hubbard, R.E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **2016**, *15*, 605–619. [CrossRef] [PubMed]
5. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* **2003**, *8*, 876–877. [CrossRef]
6. 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 settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3–25.1. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26. [PubMed]
7. 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. [CrossRef]
8. Hopkins, A.L.; Groom, C.R.; Alex, A. Ligand efficiency: A useful metric for lead selection. *Drug Discov. Today* **2004**, *9*, 430–431. [CrossRef]
9. Murray, C.W.; Rees, D.C. The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1*, 187–192. [CrossRef]
10. Romasanta, A.K.S.; van der Sijde, P.; Hellsten, I.; Hubbard, R.E.; Keseru, G.M.; van Muijlwijk-Koezen, J.; de Esch, I.J.P. When fragments link: a bibliometric perspective on the development of fragment-based drug discovery. *Drug Discov. Today* **2018**, *23*, 1596–1609. [CrossRef]
11. Tsai, J.; Lee, J.T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N.K.; et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3041–3046. [CrossRef] [PubMed]
12. Bollag, G.; Hirth, P.; Tsai, J.; Zhang, J.; Ibrahim, P.N.; Cho, H.; Spevak, W.; Zhang, C.; Zhang, Y.; Habets, G.; et al. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature* **2010**, *467*, 596–599. [CrossRef]
13. Souers, A.J.; Leverson, J.D.; Boghaert, E.R.; Ackler, S.L.; Catron, N.D.; Chen, J.; Dayton, B.D.; Ding, H.; Enschede, S.H.; Fairbrother, W.J.; et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202–208. [CrossRef]
14. Perera, T.P.S.; Jovcheva, E.; Mevellec, L.; Vialard, J.; Lange, D.D.; Verhulst, T.; Paulussen, C.; Ven, K.V.D.; King, P.; Freyne, E.; et al. Discovery and Pharmacological Characterization of JNJ-42756493 (Erdafitinib), a Functionally Selective Small-Molecule FGFR Family Inhibitor. *Mol. Cancer Ther.* **2017**, *16*, 1010–1020. [CrossRef]
15. Practical Fragments. Health and Environmental Effects of Particulate Matter (PM). 2014. Available online: http://practicalfragments.blogspot.hu/2016/10/poll-results-affiliation-metrics-and.html (accessed on 18 May 2019).
16. Practical Fragments. Poll Results: The Modern Fragment Library. Available online: http://practicalfragments.blogspot.com/2018/12/poll-results-library-vendors.html (accessed on 18 May 2019).
17. Messick, T.E.; Smith, G.R.; Soldan, S.S.; McDonnell, M.E.; Deakyne, J.S.; Malecka, K.A.; Tolvinski, L.; van den Heuvel, A.P.J.; Gu, B.-W.; Cassel, J.A.; et al. Structure-based design of small-molecule inhibitors of EBNA1 DNA binding blocks Epstein-Barr virus latent infection and tumor growth. *Sci. Transl. Med.* **2019**, *11*, eaau5612. [CrossRef] [PubMed]
18. Böttcher, J.; Dilworth, D.; Reiser, U.; Neumüller, R.A.; Schleicher, M.; Petronczki, M.; Zeeb, M.; Mischerikow, N.; Allali-Hassani, A.; Szewczyk, M.M.; et al. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat. Chem. Biol.* **2019**, *15*, 822–829. [CrossRef]
19. Keserű, G.M.; Erlanson, D.A.; Ferenczy, G.G.; Hann, M.M.; Murray, C.W.; Pickett, S.D. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **2016**, *18*, 8189–8206. [CrossRef]
20. Chen, I.-J.; Hubbard, R.E. Lessons for fragment library design: analysis of output from multiple screening campaigns. *J. Comput. Aided Mol. Des.* **2009**, *23*, 603–620. [CrossRef]

21. Dixon, S.L.; Villar, H.O. Bioactive Diversity and Screening Library Selection via Affinity Fingerprinting. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1192–1203. [CrossRef] [PubMed]
22. Roth, H.-J. There is no such thing as 'diversity'! *Curr. Opin. Chem. Biol.* **2005**, *9*, 293–295. [CrossRef] [PubMed]
23. Koutsoukas, A.; Paricharak, S.; Galloway, W.R.J.D.; Spring, D.R.; IJzerman, A.P.; Glen, R.C.; Marcus, D.; Bender, A. How Diverse Are Diversity Assessment Methods? A Comparative Analysis and Benchmarking of Molecular Descriptor Space. *J. Chem. Inf. Model.* **2014**, *54*, 230–242. [CrossRef] [PubMed]
24. Kauvar, L.M.; Higgins, D.L.; Villar, H.O.; Sportsman, J.R.; Engqvist-Goldstein, Å.; Bukar, R.; Bauer, K.E.; Dilley, H.; Rocke, D.M. Predicting ligand binding to proteins by affinity fingerprinting. *Chem. Biol.* **1995**, *2*, 107–118. [CrossRef]
25. Wawer, M.J.; Li, K.; Gustafsdottir, S.M.; Ljosa, V.; Bodycombe, N.E.; Marton, M.A.; Sokolnicki, K.L.; Bray, M.-A.; Kemp, M.M.; Winchester, E.; et al. Toward performance-diverse small-molecule libraries for cell-based phenotypic screening using multiplexed high-dimensional profiling. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10911–10916. [CrossRef] [PubMed]
26. Carbó, R.; Leyda, L.; Arnau, M. How similar is a molecule to another? An electron density measure of similarity between two molecular structures. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189. [CrossRef]
27. Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50*, 742–754. [CrossRef] [PubMed]
28. Duan, J.; Dixon, S.L.; Lowrie, J.F.; Sherman, W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* **2010**, *29*, 157–170. [CrossRef]
29. Gillet, V.J. Diversity selection algorithms. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 580–589. [CrossRef]
30. Jaccard, P. Distribution de la Flore Alpine dans le Bassin des Dranses et dans quelques régions voisines. *Bulletin de la Societe Vaudoise des Sciences Naturelles* **1901**, *37*, 241–272.
31. Tanimoto, T.T. *An Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corporation: Armonk, NY, USA, 1958.
32. Martin, E.J.; Blaney, J.M.; Siani, M.A.; Spellmeyer, D.C.; Wong, A.K.; Moos, W.H. Measuring Diversity: Experimental Design of Combinatorial Libraries for Drug Discovery. *J. Med. Chem.* **1995**, *38*, 1431–1436. [CrossRef] [PubMed]
33. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423. [CrossRef]
34. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [CrossRef]
35. Hill, M.O. Diversity and Evenness: A Unifying Notation and Its Consequences. *Ecology* **1973**, *54*, 427–432. [CrossRef]
36. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
37. Sylvester, J.J. XIX. A demonstration of the theorem that every homogeneous quadratic polynomial is reducible by real orthogonal substitutions to the form of a sum of positive and negative squares. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1852**, *4*, 138–142. [CrossRef]
38. Sauer, W.H.B.; Schwarz, M.K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987–1003. [CrossRef]
39. Jordan, J.B.; Poppe, L.; Xia, X.; Cheng, A.C.; Sun, Y.; Michelsen, K.; Eastwood, H.; Schnier, P.D.; Nixey, T.; Zhong, W. Fragment Based Drug Discovery: Practical Implementation Based on 19F NMR Spectroscopy. *J. Med. Chem.* **2012**, *55*, 678–687. [CrossRef]
40. Vulpetti, A.; Dalvit, C. Design and Generation of Highly Diverse Fluorinated Fragment Libraries and their Efficient Screening with Improved 19|NMR Methodology. *ChemMedChem* **2013**, *8*, 2057–2069. [CrossRef]
41. Visini, R.; Awale, M.; Reymond, J.-L. Fragment Database FDB-17. *J. Chem. Inf. Model.* **2017**, *57*, 700–709. [CrossRef]
42. Downs, G.M.; Barnard, J.M. Clustering Methods and Their Uses in Computational Chemistry. In *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd.: Marblehead, MA, USA, 2003; pp. 1–40. ISBN 978-0-471-43351-4.
43. Gobbi, A.; Lee, M.-L. DISE: Directed Sphere Exclusion. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 317–323. [CrossRef] [PubMed]

44. Sterling, T.; Irwin, J.J. ZINC 15—Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337. [CrossRef] [PubMed]
45. Gillet, V.J.; Willett, P.; Bradshaw, J. Identification of Biological Activity Profiles Using Substructural Analysis and Genetic Algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165–179. [CrossRef] [PubMed]
46. Weininger, D.; Weininger, A.; Weininger, J.L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101. [CrossRef]

Sample Availability: Not available.