Molecular Filters in Medicinal Chemistry

Sebastjan Kralj [1], Marko Jukič [1,2,*] and Urban Bren [1,2,3,*]

1   Laboratory of Physical Chemistry and Chemical Thermodynamics, Faculty of Chemistry
    and Chemical Engineering, University of Maribor, Smetanova 17, SI-2000 Maribor, Slovenia
2   Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska,
    Glagoljaška 8, SI-6000 Koper, Slovenia
3   Institute of Environmental Protection and Sensors, Beloruska Ulica 7, SI-2000 Maribor, Slovenia
*   Correspondence: marko.jukic@um.si (M.J.); urban.bren@um.si (U.B.); Tel.: +386-2-2294-428 (M.J.);
    +386-2-2294-421 (U.B.)

Definition: Efficient chemical library design for high-throughput virtual screening and drug design requires a pre-screening filter pipeline capable of labeling aggregators, pan-assay interference compounds (PAINS), and rapid elimination of swill (REOS); identifying or excluding covalent binders; flagging moieties with specific bio-evaluation data; and incorporating physicochemical and pharmacokinetic properties early in the design without compromising the diversity of chemical moieties present in the library. This adaptation of the chemical space results in greater enrichment of hit lists, identified compounds with greater potential for further optimization, and efficient use of computational time. A number of medicinal chemistry filters have been implemented in the Konstanz Information Miner (KNIME) software and analyzed their impact on testing representative libraries with chemoinformatic analysis. It was found that the analyzed filters can effectively tailor chemical libraries to a lead-like chemical space, identify protein–protein inhibitor-like compounds, prioritize oral bioavailability, identify drug-like compounds, and effectively label unwanted scaffolds or functional groups. However, one should be cautious in their application and carefully study the chemical space suitable for the target and general medicinal chemistry campaign, and review passed and labeled compounds before taking further in silico steps.

1. Introduction

Physical screening of large libraries was the predominant method for initial steps in drug discovery in the past, and is now effectively complemented by in silico counterpart, namely, HTVS or high-throughput virtual screening [1]. Successful virtual screening campaigns can achieve high confirmed hit rates and such methods are gaining in strength with hardware and software development [2]. The use of virtual screening on virtual compound libraries reduces the number of potential lead molecules to be evaluated in vitro, increasing time- and cost-efficiency of the drug development process [3]. The use of virtual compound libraries brings with it the vast expansion of chemical space that can be searched [4]. If conventional physical libraries of pharmaceutical companies are on the scale of 10^6 to 10^7 compounds, advanced virtual compound libraries such as GDB-17 can reach up to 1.16^{11} [5–8]. Despite its sheer size, GDB-17 consists only of molecules with up to 17 atoms of C, N, O, S, and halogens, which points to the fact that the number of possible unique organic molecules is immense, with estimates from 10^{13} to 10^{180}, depending on inclusion criteria [8,9]. Even with the rapid development of both computational power in the form of super-computers (HPC) and advances in the methods used, it is impossible to search such vast chemical spaces [3]. Such libraries force medicinal chemists to make the trade-off between completeness and screenability, as complete libraries are not easily

screened, but screenable libraries are not complete, and could perhaps cover only specific chemical spaces [3].

Methods such as molecular docking used for lead identification and molecular dynamics simulations for lead optimization require vast libraries to be processed and focused, as time consumption associated with such methods is far greater than that of simple two-dimensional methods [10]. With computer-aided drug design (CADD) the general workflow follows three steps: (1) filtering of large compound libraries into focused libraries based on the users need, (2) discovery and optimization of lead compounds, and (3) development of novel compounds, with steps 2 and 3 repeated until compounds with desired properties are obtained [11,12]. Since hit rates of screening campaigns are on average as low as 1%, the most efficient and quick way to increase hit-rates is to use molecular filters (or cheminformatics filters) [13]. Molecular filters narrow down the chemical space of large libraries towards predetermined goals by removing unwanted chemical structures and properties, with the majority of the filters developed focusing the libraries towards drug-like and bioavailable molecules (Figure 1) [4,14]. Pioneered by prolific Chris Lipinski and coworkers, molecular filters were developed by intelligent analysis of drug hits obtained in Pfizer's laboratories, with the assumption that poor physio-chemical properties predominate in many compounds that enter but fail during pre-clinical stages and Phase 1 safety evaluations. By analyzing data from 2245 compounds, they were able to determine molecular features shared among orally available drugs, that critically influence pharmacokinetics [14,15]. The term drug-likeness, often associated with the use of filters and used in different ways by different authors, generally refers to compounds with desirable properties, such as oral bioavailability, low toxicity, suitable clearance rate, and membrane permeability, which are properties often found in the majority of approved drugs [14,16–18]. An alternative for narrowing the chemical space of compound libraries is clustering, an approach based on the premise that similar compounds have similar activity. Unlike molecular filtering, clustering works in a less focused way, as compounds are separated based on similarity with a selection of representative compounds from subsequent groups [19]. However, unlike molecular filtering, where library size does not impact the choice of the filter used, the choice of the clustering approach is library size dependent. Hierarchical clustering is preferred for small libraries and faster non-hierarchical clustering is preferred for large libraries [19,20].
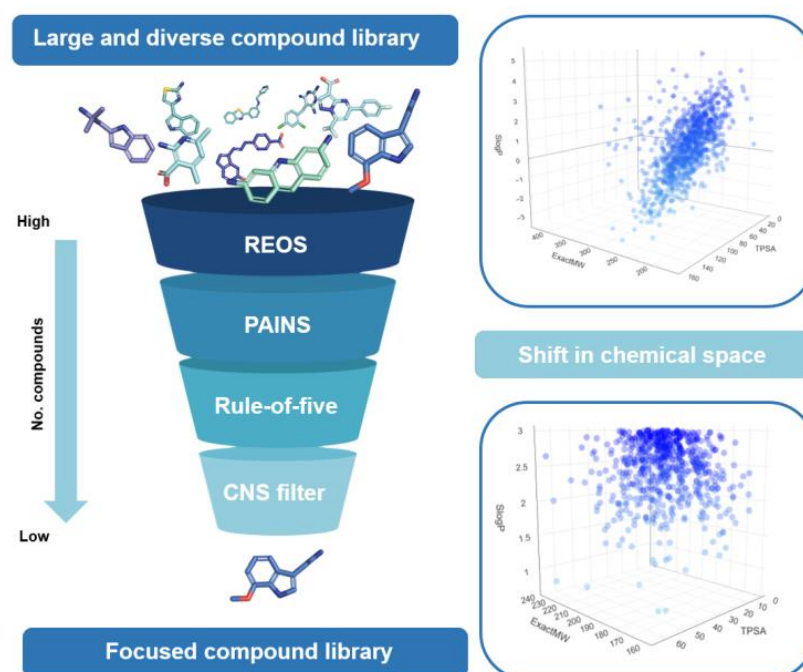

Figure 1. Impact of filtering on the size and chemical space of a large and diverse library.

2. Types of Filters

Compound filters in use today can roughly be split into two groups: filters that exclude compounds based on the presence of functional groups and filters that exclude compounds based on certain descriptor properties. The first group of filters is, therefore, named functional group filters, and the second group is named property filters [12].

2.1. Functional Group Filters

Functional group filters are based on the premise that covalent chemistry is undesired in drug design and filters out electrophilic functional groups, while some of the filters focus on removing optically interfering components, aggregators, fluorescent compounds, etc. [18,21,22]. Compounds with aforementioned functionality are known to interfere with screening tests, often appearing as false positives in HTS screening scenarios [23]. The main advantage of functional filters is the removal of compounds that would increase expenses of assay in vitro screening. However, the downside is the removal of potential covalent drug candidates and should be used with care.

Rapid elimination of swill (REOS) is a functional filter, which was, at the time of development by researchers at Vertex pharmaceuticals, the first of its kind [24]. The filter is based on 117 SMARTS strings collected from the literature data describing non-druglike functionalities associated with promiscuous ligands and frequent hitters, such as reactive moieties and known toxicophores. The concept and main advantage of this filter is to increase screening efficiency through identification and elimination of compounds that are not worthy of serious consideration as lead-like compounds. Several other efforts at developing such functional filters have been made, most notably by groups from Amgen [25], University of New Mexico [26], and Eli Lilly [27].

Pan-assay interference compounds (PAINS) is a functional group filter that applies the same approach of targeting frequent hitters (promiscuous compounds) as the REOS filter. It does so by using a list of 480 functional groups shared by many PAINS and then comparing them to the input database. Compounds that possess undesired functional groups are flagged and can be filtered out [22]. Some examples of such problematic sub-structures are quinones, rhodanines, toxoflavins, curcumin, 2-aminothiophenes, etc.

The aggregators filter is based on a combined approach of using lipophilicity, affinity, and similarity to a database of known aggregators in order to determine the propensity of organic compounds for colloidal aggregation. This filter is, in essence, a hybrid filter as it combines a functional group filter, where input molecules are compared to a database of known aggregators using the Tanimoto coefficient, and a property filter, where SlogP descriptor cut-off of <3 is set. The SlogP cut-off was set based on the fact that 80% of known aggregators surpassed the set value [28].

2.2. Property Filters

Physio-chemical property filters predominantly aim at addressing ADMET (absorption, distribution, metabolism, excretion, toxicity) issues that may arise in downstream drug-development process [29]. The knowledge-based approach of developing such filters is based on the fact that certain descriptors such as logP, molecular weight (MW), and number of hydrogen bond acceptors/donors have been correlated with oral bioavailability [14,17]. Such information can be used to derive descriptors cut-off limits that bias the chemical space of a library towards the drug-like paradigm [4]. The key advantage of using property filters in drug discovery is the shift of the general chemical space towards the desired chemical space. As similar compounds occupy a similar chemical space and similar compounds usually have similar activity, focusing the chemical space through filtering should, in theory, increase the chances of finding prospective drugs [19]. The main downside of such filters is the chemical space bias, as the knowledge-based approach used for filter design will always shift the chemical space towards the same paradigm, eliminating diverse chemical entities. This points to the fact that property filters should be

dynamic in nature and that their use should be regarded more as a guideline rather than a strict rule.

2.2.1. Bioavailability

Lipinski's rule of 5, one of the fundamental chemo-informatics filters, represents a staple among property filters. The filter focuses the chemical space towards the drug-like narrative and ADMET issue aversion through a set of rules (molecular weight (MW) $\leq$ 500 Da, logP $\leq$ 5, hydrogen bond donor (HBD) $\leq$ 5, hydrogen bond acceptor (HBA) $\leq$ 10. The rules were derived from a subset of 2245 drugs from the World Drug Index [14,30]. The filter helps to predict if a biologically active molecule is likely to have the chemical and physical properties to be orally bioavailable. However, as with all filters, it should be applied with caution [31].

The Veber filter was created by analyzing bioavailability data of over 1100 drug candidates processed at GlaxoSmithKline. The filter contains two simple rules (total polar surface area (TPSA) $\leq$ 140 Å^2, rotatable bonds $\leq$ 10) that compounds should adhere to for optimized bioavailability [32].

The Egan filter consists of a set of rules (logP $\leq$ 5.88, TPSA $\leq$ 131.6 Å^2) determined by using multivariate statistics on data of compounds both well and poorly absorbed in humans. Only two descriptors (logP and TPSA) were chosen for inclusion when determining membrane permeability, with the goal of good bioavailability [33]. With bioavailability filters, it is important to note that they can remove compounds that pass into cells with carrier-mediated transport and active-uptake [34].

The Palm filter was designed based on evaluation of dynamic surface properties of drug molecules and drug absorption in two cell models. The results show that polar surface area is a better descriptor for intestinal drug absorption than logP. The findings were confirmed with 20 model drugs, that had various absorption rates. The cut-off for the standard orally bioavailable drugs was determined at TPSA < 140 Å^2 and at TPSA < 63 Å^2 for the enhanced version filtering for strictly orally bioavailable molecules [35,36].

2.2.2. Drug-Likeness

The Mozziconacci filter builds upon the foundation of the rule of 5, and was designed by analyzing 15 different commercial and freely available chemical libraries for drug-likeness (number of halogen atoms $\leq$ 7, number of nitrogens $\geq$ 1, number of oxygens $\geq$ 1, number of rings $\leq$ 6, rotatable bonds $\leq$ 15) [37]. The REOS filter is a hybrid method that combines a set of functional group filters described above and a set of property filters. This part of the filter is useful for determining drug-like molecules that will later be passed through the functional filter. It involves several criteria for the drug to meet such as 200 < MW < 500, -5 < logP < 5, HBD < 5, HBA <10, -2 < formal Charge < 2, number of rotatable bonds < 8, and 15 < number of heavy atoms < 50 [21]. The Ghose filter is a knowledge-based filter which aims to provide a user with a quantitative and qualitative representation of drug-like chemical space that can be used for designing combinatorial or medicinal chemistry libraries for drug discovery. The rules defining this filter were derived by analyzing drug databases and are as follows: 160 < MW < 480, -0.4 < logP < 5.6, 20 < number of atoms < 70 and 40 < molar refractivity < 130 [38]. Oprea et al. have devised a filter to remove property extrema of unwanted properties. By examining property distribution in several databases and using Pareto analyses, drug-like properties of compounds were determined. Such compounds adhere to the following rules: HBD < 2, 2 < HBA < 10, 2 < rotatable bonds < 8 and 1 < number of rings < 4. The authors emphasize that such filters do not remove reactive species, pointing to the fact that the use of several filters is optimal for library design [39].

2.2.3. Lead-Likeness

The rule of 3 expanded on the findings of some authors that lead-like compounds exhibit less complexity when compared to drugs, and often have lipophilicity and MW

increased during optimization. Such optimization often means losing compliance with traditional rules such as rule of 5 [40]. The rule of 3 uses four rules (MW $\leq$ 300 Da, logP $\leq$ 3, HBD $\leq$ 3, HBA $\leq$ 3) and has been optimized to define lead like compounds for further fragment-based design [41].

2.2.4. Central Nervous System Activity (Blood–Brain Barrier Permeability)

Besides bioavailability and drug-likeness, several filters for passing the blood–brain barrier have been developed as well. Such filters are important both for the development of peripherally selective drugs or CNS-active drugs. With peripherally selective drugs, the passage of drugs into the CNS is undesired as it may lead to the occurrence of various side effects [42]. As blood–brain barrier permeability filters can include or exclude compounds for both CNS and non-CNS drug-development cases, they are useful in virtually every drug-development filtering process [43]. The Van der Waterbeemd filter was one of the first filters designed for estimation of blood–brain barrier crossing with two rules (MW $\leq$ 450 Da, TPSA $\leq$ 90 Å^2) which were derived by examination of lipophilicity, H-bonding capacity, and molecular shape of 125 marketed CNS and CNS-inactive drugs [44,45]. The other filter for CNS activity is the Murcko filter with five rules (200 $\leq$ MW $\leq$ 400, LogP $\leq$ 5.2, HBA $\leq$ 4, HBD $\leq$ 3, number of rotatable bonds $\leq$ 7) that aim to focus libraries for CNS activity [46]. Modern approaches for blood brain permeability prediction are algorithms based on calculation of physiochemical descriptors, one such example is the "BBB score" designed by Gupta et al. which uses five descriptors (no. atomic rings, no. heavy atoms, MW, HBD, and HBA) and represents a useful addition to the blood–brain permeability prediction arsenal of medicinal chemists [47].

2.2.5. Protein–Protein Interaction Inhibitors

Rule of 4 is a set of rules for identification of druggable protein–protein-interaction (PPI) inhibitors (MW $\geq$ 400 Da, logP $\geq$ 4, number of rings (NoR) $\geq$ 4, HBA $\geq$ 4). As protein–protein inhibitors are often of large molecular weight and possess many hydrogen-bond acceptors, the rules defining this filter deviate from the drug-like paradigm and were derived from analyzing the 2P2I database which contains protein–protein interaction inhibitors. The filter focuses the chemical space towards larger molecules capable of forming several interactions [48]. Just like the blood brain permeability filters, the PPI filter represents a specific filter, and its logic can be reversed, as some of the properties desired with PPI inhibitors, such as large molecular weight and large number of HBA, are the opposite to those found in traditional drug-like filters. This dual nature of specific filters is a key advantage over general filters.

3. Limitations of Filter Use

Although molecular filters have firmly established themselves as useful tools in drug discovery, with filters such as Lipinski's rule of 5, the Ghose filter, and the Veber filter frequently employed for compound filtering, there remains a fair amount of criticism from experts associated with inconsiderate use [14,32,38]. Some claim that compound filters are overzealous and may lead to the elimination of potentially valuable therapeutics, as is the case in the article by Tropsha et al. that questions the guidelines for use of the PAINS filter [49]. Since filters work on a pass or fail basis, they often ignore exceptions where the majority of filters would fail as exemplified by cyclosporine and erythromycin and compounds that are substrates for drug transporters. Recent work also suggests that carrier-mediated transport and active-uptake may be more common than assumed [18,34]. The use of filters always carries an informed decision made by the user, as they define favorable and undesirable features of molecules to be filtered [50]. Such a decision should always be taken considering the suitable biological context and filters used with informed care [4,51]. Although filters demonstrate low accuracy with regard to passing/failing registered drugs, the data seem to be pointing to the direction of filters possessing great value as tools for early drug research and can increase return on computational investment.

Alongside the question of the choice of filter, the question of when to use the compound filter is important as well. The common practice is to incorporate filters upfront, as using the computationally undemanding methods first and computationally more demanding methods second makes sense from the perspective of return-on-computational-investment [52]. Additionally, the use of filters upfront is backed up by research indicating that lead-like compounds exhibit less molecular complexity and are less hydrophobic when compared to approved drugs, and, as such, optimization of simple leads into drugs is favored in the drug-design process [40]. Another great benefit of using compound filters is to use them in conjunction, as the users can filter out problematic functional groups that appear as frequent hitters and later focus the library, for instance, towards CNS-active compounds if this is what the biological context demands. With regard to optimization of the workflow for speed when using the approach of consecutive filtering, it is advised that the user to first use simple property filters (e.g., Lipinski) which work faster than functional filters that often require sub-structure counting (e.g., REOS). In this way, the more time-consuming filters are used on smaller libraries.

4. Impact on Chemical Space

In order to test the impact of molecular filters on chemical space of large drug-like libraries, a set of molecular filters was tested by filtering two compound subsets and evaluating the impact. The filters used are implemented in the form of a KNIME workflow, and its design, implementation, and testing is described in detail in previously published work (available at: https://hub.knime.com/-/spaces/-/latest/~xoK5FQgB_5Jmg54V/, accessed on 15 March 2023) [4]. KNIME (version 4.7) is an open-source data analytics and integration platform that uses the concept of a graphical user interface and modular data pipelining to create an intuitive environment for complex data processing tasks. Along with many custom nodes developed for the pharmaceutical industry, it supports large-scale HTVS efforts through its KNIME server implementation, making it a perfect tool for early drug-discovery processes. The sample groups of compounds that were filtered were obtained by random sampling the GDB-17 library and the ZINC in-stock library [8,53,54]. The two samples of 100,000 compounds were obtained using the row sampling node, with sampling set to random. The statistic node was used to calculate the number of passed compounds and descriptor values.

The retention rate of the filtered libraries gives an insight into filter strictness (Figure 2). Of note is that GDB-17 only contains compounds with a length of up to 17 atoms as opposed to ZINC. For further analysis of chemical space during the filtering process, the ZINC in-stock library was used. Such diverse chemical space is often present with compound libraries that one might encounter before the filtering process. The scatterplot analysis shows a distinct move for the lead-like compounds from the unfiltered green compounds and the drug-like compounds, supporting the narrative that lead-like molecules exhibit less complexity (Figure 3). The unfiltered compounds naturally have greater outliers as there are no property restrictions. The majority of the chemical space for Lipinski and Oprea filters overlap and show that drug-likeness and bioavailability are very similar, also putting these filters in the same category as expected.

With specific filters, such as the rule of 4 for PPI inhibitors and Van de Waterbeemd for passage of the blood–brain barrier, it can been seen sharp shifts in chemical space (Figure 4). The chemical space of molecules that passed the rule of 4 filter is shifted towards large molecules with large TPSA and SlogP values, while the chemical space of Van de Waterbeemd molecules is shifted towards smaller molecules. It is also interesting to see that functional group filters such as REOS do not change the chemical space as the overlap with the unfiltered group is large, despite the REOS functional filter having a retention rate of 68%. The filter is aimed towards targeting reactive species which are in general present regardless of the physio-chemical properties of compounds.
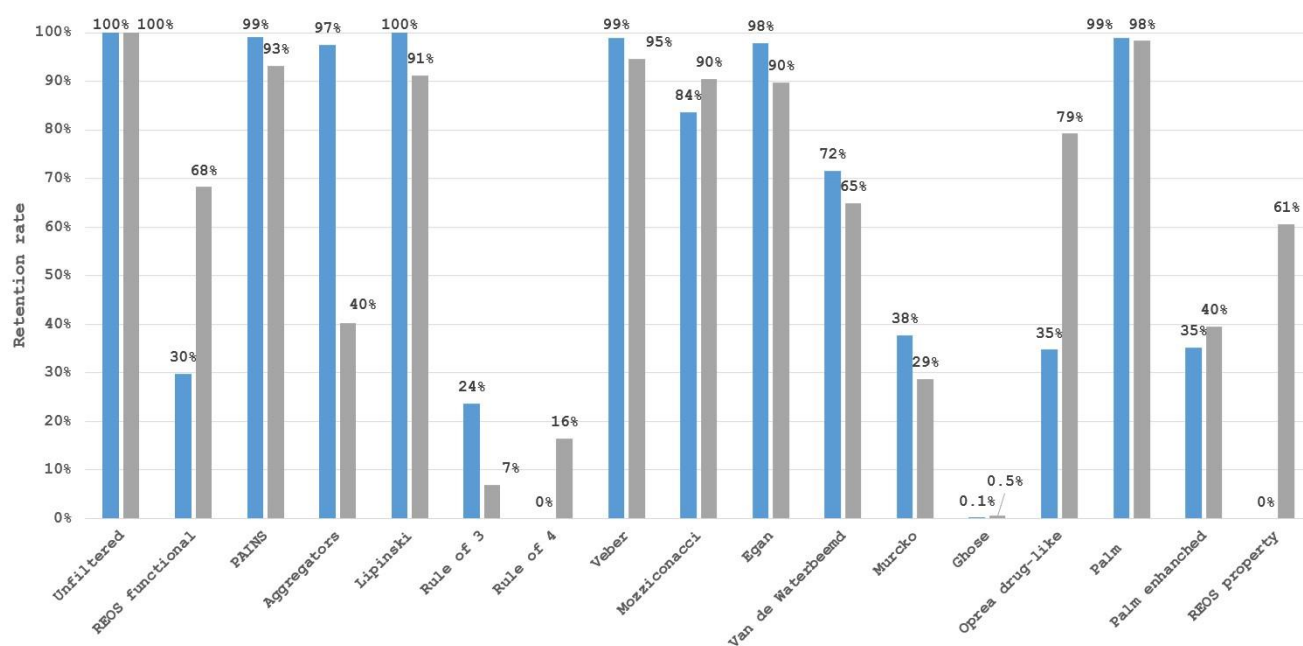
Figure 2. Retention rate for individual filters used for GDB-17 (blue) and ZINC (grey).
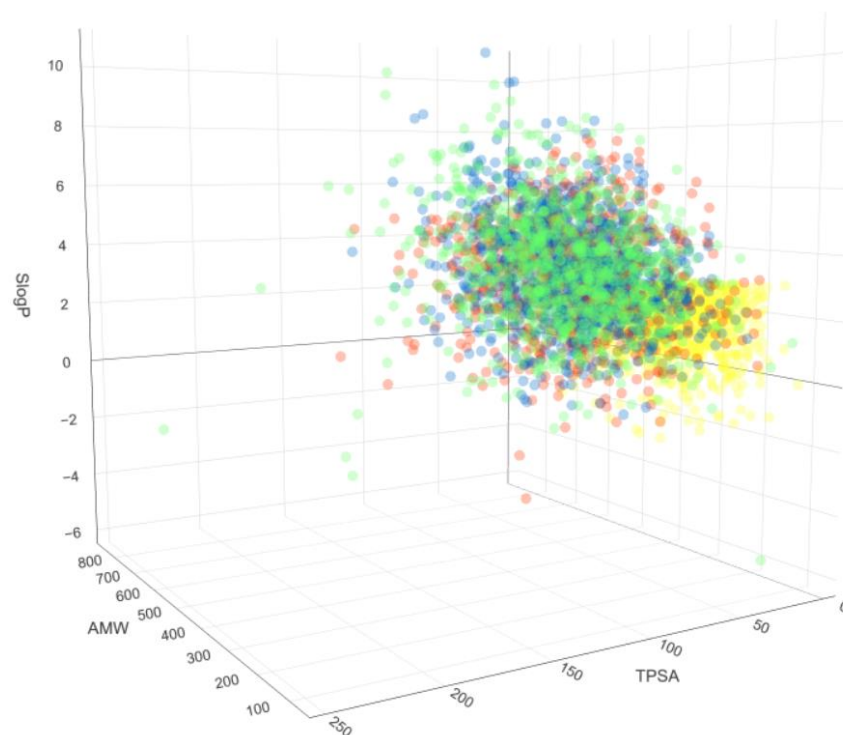

Figure 3. Three-dimensional scatter plot of AMW, TPSA, and SlogP for compounds filtered by Oprea (blue), Lipinski (red), and rule of 3 (yellow) filters and the unfiltered library (green).
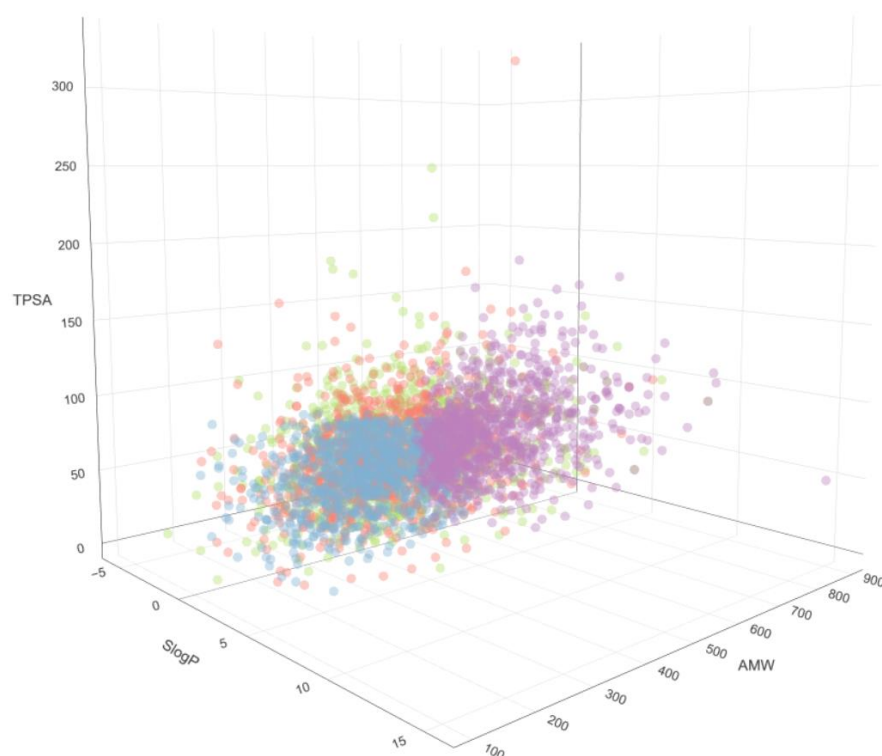
Figure 4. Three-dimensional scatterplot of AMW, TPSA, and SlogP for compounds filtered by the REOS functional (red), Van de Waterbeemd (blue), and rule of 4 filters (purple) and the unfiltered library (green).

5. Conclusions

The use of filters in molecular screening campaigns is of vital importance. With the development of methods and the fast improvement of computational power, virtual screening in drug design (HTVS) and CADD in general are gaining power. However, the systematic coverage of complete organic small-molecule chemical space is still above the capabilities of a typical HPC-supported in silico pipeline. Novel generative models are capable of producing vast libraries, and their usefulness is in active evaluation. Currently, filters are applied to compound libraries before screening with computationally intensive methods. As more computationally intensive methods are applied to fewer compounds, this strategy increases the likelihood of finding matches with the desired features and increases the return on computational investment. One such successful example of the use of molecular filters is the work of Jukič et al., where pre-filtering of the library enabled extensive screening using computationally intensive methods, such as linear interaction energy calculations [10]. The herein mentioned filters were applied sequentially to a collection of commercial libraries. The filtering included a coarse pre-filter for large molecules, small fragments, and aggregators, followed by a PAINS and REOS filter. An important concept of expanding the chemical space of the final library by generating structural isomers of the filtered library was highlighted. A similar example of the successful application of filters in drug development was provided by Kolarič et al. [48]. To narrow down the large library, the authors applied the functional REOS, PAINS, and aggregators filters before performing a final filtering of the chemical library. The inhibitory potential of the small molecules herein discovered was confirmed in vitro [49]. The use of filters is almost universal in the virtual drug discovery programs, with an interesting recent application by Shoichet et al., where they employed typical filtering approaches such as Ro5 to analyze various chemical libraries and examine the effect of library size on the chemical space [55]. The same group wonderfully summarized the effective use of filtering in a practical guide to virtual screening [56]. In addition to filtering in-house assembled libraries, it also supports the use

of filtering pipeline on custom commercial compound collections that are popular with online drug vendors. Namely, custom libraries offered by vendors often contain unwanted compounds or lack information on library construction [12]. The filtering pipeline can be constructed by referencing the primary literature or employing various implementations as seen in bioinformatics packages. One such example is implementation in the freely available KNIME software (https://gitlab.com/Jukic/knime_medchem_filters or alternatively at https://hub.knime.com/marko_/spaces/Medicinal%20Chemistry%20Filtering/latest/~XQX98NGeZEgCxQ_j/; accessed on 18 April 2023) [4]. Such implementations allow the user to participate in further improvements of the software, and the users are encouraged to comment and file issues so that together a current and useful filtering pipeline can be constructed. Since filters for medicinal chemistry are a knowledge-based approach, such widespread use of open-access software could lead to an expansion towards specific chemistry filters (e.g., rule-of-4). Another major advantage of this approach is the ability to monitor flagged compounds at each step, an important aspect in terms of filter stringency. Overall, an efficient filtering pipeline can offer an effective preparation of tailored chemical libraries and help in novel drug design.

References

1. Shoichet, B.K. Virtual Screening of Chemical Libraries. *Nature* **2004**, *432*, 862–865. [CrossRef] [PubMed]
2. Doman, T.N.; McGovern, S.L.; Witherbee, B.J.; Kasten, T.P.; Kurumbail, R.; Stallings, W.C.; Connolly, D.T.; Shoichet, B.K. Molecular Docking and High-Throughput Screening for Novel Inhibitors of Protein Tyrosine Phosphatase-1B. *J. Med. Chem.* **2002**, *45*, 2213–2221. [CrossRef] [PubMed]
3. Van Hilten, N.; Chevillard, F.; Kolb, P. Virtual Compound Libraries in Computer-Assisted Drug Discovery. *J. Chem. Inf. Model.* **2019**, *59*, 644–651. [CrossRef] [PubMed]
4. Kralj, S.; Jukič, M.; Bren, U. Comparative Analyses of Medicinal Chemistry and Cheminformatics Filters with Accessible Implementation in Konstanz Information Miner (KNIME). *Int. J. Mol. Sci.* **2022**, *23*, 5727. [CrossRef] [PubMed]
5. Blay, V.; Tolani, B.; Ho, S.P.; Arkin, M.R. High-Throughput Screening: Today's Biochemical and Cell-Based Approaches. *Drug Discov. Today* **2020**, *25*, 1807–1821. [CrossRef]
6. Bakken, G.A.; Bell, A.S.; Boehm, M.; Everett, J.R.; Gonzales, R.; Hepworth, D.; Klug-McLeod, J.L.; Lanfear, J.; Loesel, J.; Mathias, J.; et al. Shaping a Screening File for Maximal Lead Discovery Efficiency and Effectiveness: Elimination of Molecular Redundancy. *J. Chem. Inf. Model.* **2012**, *52*, 2937–2949. [CrossRef]
7. Njoroge, M.; Njuguna, N.M.; Mutai, P.; Ongarora, D.S.B.; Smith, P.W.; Chibale, K. Recent Approaches to Chemical Discovery and Development against Malaria and the Neglected Tropical Diseases Human African Trypanosomiasis and Schistosomiasis. *Chem. Rev.* **2014**, *114*, 11138–11163. [CrossRef]
8. Ruddigkeit, L.; van Deursen, R.; Blum, L.C.; Reymond, J.-L. Enumeration of 166 Billion Organic Small Molecules in the Chemical Universe Database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875. [CrossRef]
9. Gorse, A.-D. Diversity in Medicinal Chemistry Space. *Curr. Top. Med. Chem.* **2006**, *6*, 3–18. [CrossRef]
10. Jukič, M.; Janežič, D.; Bren, U. Ensemble Docking Coupled to Linear Interaction Energy Calculations for Identification of Coronavirus Main Protease (3CLpro) Non-Covalent Small-Molecule Inhibitors. *Molecules* **2020**, *25*, 5808. [CrossRef]

11. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E.W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395. [CrossRef] [PubMed]
12. Kralj, S.; Jukič, M.; Bren, U. Commercial SARS-CoV-2 Targeted, Protease Inhibitor Focused and Protein–Protein Interaction Inhibitor Focused Molecular Libraries for Virtual Screening and Drug Design. *Int. J. Mol. Sci.* **2021**, *23*, 393. [CrossRef] [PubMed]
13. Thorpe, D.S.; Edith Chan, A.W.; Binnie, A.; Chen, L.C.; Robinson, A.; Spoonamore, J.; Rodwell, D.; Wade, S.; Wilson, S.; Ackerman-Berrier, M.; et al. Efficient Discovery of Inhibitory Ligands for Diverse Targets from a Small Combinatorial Chemical Library of Chimeric Molecules. *Biochem. Biophys. Res. Commun.* **1999**, *266*, 62–65. [CrossRef]
14. Lipinski, C.A. Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235–249. [CrossRef] [PubMed]
15. Oprea, T. Virtual Screening in Lead Discovery: A Viewpoint. *Molecules* **2002**, *7*, 51–62. [CrossRef]
16. Muegge, I. Pharmacophore Features of Potential Drugs. *Chem. Weinh. Bergstr. Ger.* **2002**, *8*, 1976–1981. [CrossRef]
17. Walters, W.P.; Murcko, A.A.; Murcko, M.A. Recognizing Molecules with Drug-like Properties. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384–387. [CrossRef]
18. Walters, W.P.; Murcko, M.A. Prediction of "Drug-Likeness". *Adv. Drug Deliv. Rev.* **2002**, *54*, 255–271. [CrossRef]
19. Lumley, J.A. Compound Selection and Filtering in Library Design. *QSAR Comb. Sci.* **2005**, *24*, 1066–1075. [CrossRef]
20. Pascual, R.; Borrell, J.I.; Teixidó, J. Analysis of Selection Methodologies for Combinatorial Library Design. *Mol. Divers.* **2000**, *6*, 121–133. [CrossRef]
21. Walters, W.P.; Namchuk, M. Designing Screens: How to Make Your Hits a Hit. *Nat. Rev. Drug Discov.* **2003**, *2*, 259–266. [CrossRef]
22. Baell, J.B.; Holloway, G.A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719–2740. [CrossRef] [PubMed]
23. Thorne, N.; Auld, D.S.; Inglese, J. Apparent Activity in High-Throughput Screening: Origins of Compound-Dependent Assay Interference. *Curr. Opin. Chem. Biol.* **2010**, *14*, 315–324. [CrossRef]
24. Walters, W.P.; Stahl, M.T.; Murcko, M.A. Virtual Screening—An Overview. *Drug Discov. Today* **1998**, *3*, 160–178. [CrossRef]
25. Rishton, G.M. Reactive Compounds and in Vitro False Positives in HTS. *Drug Discov. Today* **1997**, *2*, 382–384. [CrossRef]
26. Yang, J.J.; Ursu, O.; Lipinski, C.A.; Sklar, L.A.; Oprea, T.I.; Bologa, C.G. Badapple: Promiscuity Patterns from Noisy Evidence. *J. Cheminform.* **2016**, *8*, 29. [CrossRef] [PubMed]
27. Bruns, R.F.; Watson, I.A. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *J. Med. Chem.* **2012**, *55*, 9763–9772. [CrossRef]
28. Irwin, J.J.; Duan, D.; Torosyan, H.; Doak, A.K.; Ziebart, K.T.; Sterling, T.; Tumanian, G.; Shoichet, B.K. An Aggregation Advisor for Ligand Discovery. *J. Med. Chem.* **2015**, *58*, 7076–7087. [CrossRef]
29. Huggins, D.J.; Venkitaraman, A.R.; Spring, D.R. Rational Methods for the Selection of Diverse Screening Compounds. *ACS Chem. Biol.* **2011**, *6*, 208–217. [CrossRef]
30. 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 Deliv. Rev.* **2001**, *46*, 3–26. [CrossRef]
31. O'Hagan, S.; Swainston, N.; Handl, J.; Kell, D.B. A 'Rule of 0.5' for the Metabolite-Likeness of Approved Pharmaceutical Drugs. *Metabolomics* **2015**, *11*, 323–339. [CrossRef] [PubMed]
32. Veber, D.F.; Johnson, S.R.; Cheng, H.-Y.; 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. [CrossRef] [PubMed]
33. Egan, W.J.; Merz, K.M.; Baldwin, J.J. Prediction of Drug Absorption Using Multivariate Statistics. *J. Med. Chem.* **2000**, *43*, 3867–3877. [CrossRef] [PubMed]
34. Dobson, P.D.; Kell, D.B. Carrier-Mediated Cellular Uptake of Pharmaceutical Drugs: An Exception or the Rule? *Nat. Rev. Drug Discov.* **2008**, *7*, 205–220. [CrossRef]
35. Palm, K.; Luthman, K.; Unge, A.-L.; Strandlund, G.; Artursson, P. Correlation of Drug Absorption with Molecular Surface Properties. *J. Pharm. Sci.* **1996**, *85*, 32–39. [CrossRef]
36. Palm, K.; Stenberg, P.; Luthman, K.; Artursson1, P. Polar Molecular Surface Properties Predict the Intestinal Absorption of Drugs in Humans. *Pharm. Res.* **1997**, *14*, 568–571. [CrossRef]
37. Morin-Allory, L.; Mozziconacci, J.C.; Arnoult, E.; Baurin, N.; Marot, C. *Preparation of a Molecular Database from a Set of 2 Million Compounds for Virtual Screening Applications: Gathering, Structural Analysis and Filtering*; Institut de Chimie Organique et Analytique, Universite d'Orleans: Orleans, France, 2003.
38. Ghose, A.K.; Viswanadhan, V.N.; Wendoloski, J.J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55–68. [CrossRef] [PubMed]
39. Oprea, T.I. Property Distribution of Drug-Related Chemical Databases. *J. Comput. Aided Mol. Des.* **2000**, *14*, 251–264. [CrossRef] [PubMed]
40. Oprea, T.I.; Davis, A.M.; Teague, S.J.; Leeson, P.D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1308–1315. [CrossRef]
41. 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]

42. Di, L.; Kerns, E.H. (Eds.) *Blood-Brain Barrier in Drug Discovery: Optimizing Brain Exposure of CNS Drugs and Minimizing Brain Side Effects for Peripheral Drugs*; Wiley: Hoboken, NJ, USA, 2015; ISBN 978-1-118-78835-6.

43. On behalf of the 2013 CINP Summit Group. Securing the Future of Drug Discovery for Central Nervous System Disorders. *Nat. Rev. Drug Discov.* **2014**, *13*, 871–872. [CrossRef] [PubMed]

44. Van de Waterbeemd, H. Physicochemical Approaches to Drug Absorption. In *Methods and Principles in Medicinal Chemistry*; van de Waterbeemd, H., Testa, B., Eds.; Wiley: Hoboken, NJ, USA, 2008; Volume 40, ISBN 978-3-527-32051-6.

45. van De Waterbeemd, H.; Camenisch, G.; Folkers, G.; Raevsky, O.A. Estimation of Caco-2 Cell Permeability Using Calculated Molecular Descriptors. *Quant. Struct.-Act. Relatsh.* **1996**, *15*, 480–490. [CrossRef]

46. Ajay; Bemis, G.W.; Murcko, M.A. Designing Libraries with CNS Activity. *J. Med. Chem.* **1999**, *42*, 4942–4951. [CrossRef] [PubMed]

47. Gupta, M.; Lee, H.J.; Barden, C.J.; Weaver, D.F. The Blood–Brain Barrier (BBB) Score. *J. Med. Chem.* **2019**, *62*, 9824–9836. [CrossRef]

48. Morelli, X.; Bourgeas, R.; Roche, P. Chemical and Structural Lessons from Recent Successes in Protein–Protein Interaction Inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15*, 475–481. [CrossRef]

49. Capuzzi, S.J.; Muratov, E.N.; Tropsha, A. Phantom PAINS: Problems with the Utility of Alerts for P an- A Ssay IN Terference Compound, S. *J. Chem. Inf. Model.* **2017**, *57*, 417–427. [CrossRef]

50. Shultz, M.D. Two Decades under the Influence of the Rule of Five and the Changing Properties of Approved Oral Drugs: Miniperspective. *J. Med. Chem.* **2019**, *62*, 1701–1714. [CrossRef]

51. Olah, M.M.; Bologa, C.G.; Oprea, T.I. Strategies for Compound Selection. *Curr. Drug Discov. Technol.* **2004**, *1*, 211–220. [CrossRef]

52. Charifson, P.S.; Walters, W.P. Filtering Databases and Chemical Libraries. *J. Comput. Aided Mol. Des.* **2002**, *16*, 311–323. [CrossRef]

53. Polishchuk, P.G.; Madzhidov, T.I.; Varnek, A. Estimation of the Size of Drug-like Chemical Space Based on GDB-17 Data. *J. Comput. Aided Mol. Des.* **2013**, *27*, 675–679. [CrossRef]

54. Irwin, J.J.; Shoichet, B.K. ZINC—A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182. [CrossRef] [PubMed]

55. Lyu, J.; Irwin, J.J.; Shoichet, B.K. Modeling the expansion of virtual screening libraries. *Nat. Chem. Biol.* **2023**, 1–7. [CrossRef] [PubMed]

56. Bender, B.J.; Gahbauer, S.; Luttens, A.; Lyu, J.; Webb, C.M.; Stein, R.M.; Fink, E.A.; Balius, T.E.; Carlsson, J.; Irwin, J.J. A practical guide to large-scale docking. *Nat. Protoc.* **2021**, *16*, 4799–4832. [CrossRef] [PubMed]