Drug-Likeness Analysis

Soumya Jyoti Sain*

Abstract

In the early stages of drug discovery, in-silico evaluation of drug-likeness is a crucial step for assessing the viability of potential drug molecules. This is achieved by implementing various predefined physicochemical guidelines, such as those proposed by Lipinski, Veber, Ghose, and Egan. This project presents a comparative analysis of these guidelines by applying them to 1037 FDA-approved drugs. Molecular structures were generated from the SMILES strings, and 8 molecular descriptors were calculated using RDKit. Each filter was applied to each molecule and classified as pass or fail based on compliance with all criteria. Following this, a comparative analysis was performed, which revealed that the Lipinski and Veber filters exhibited similar permissiveness, allowing over 900 molecules to pass, whereas the Egan filter was slightly more stringent and the Ghose filter was substantially more restrictive, with most drug molecules failing the filter. Distribution analysis showed that most drugs passed 3 filters, followed by those passing all 4. Additionally, the descriptor violations responsible for failures in the Lipinski and Ghose filters were identified and analysed. This study overall provides insights on the viability of computational filters.

1 Introduction

Physicochemical filters are integral to medicinal chemistry and drug discovery, with their computational nature reducing the need for resource-intensive early experimental screening. The project employs a comparative study of drug-likeness on FDA-approved drugs. By examining compliance patterns, the study highlights that these criteria serve as advisory guidelines rather than absolute requirements.

2 Dataset

The dataset was obtained from PubChem, containing SMILES strings and various molecular descriptors. For consistency purposes, the SMILES were first converted into RDKit-compatible molecules, and then the original molecular descriptors were replaced with RDKit molecular descriptors.

*Code Available at https://github.com/DevSoumyaJ/drug-likeness-analysis/

3 Methodology

3.1 Descriptor Calculation

Table 1: Molecular descriptors used in the study

Descriptor	Description
Molecular Weight	Molecular mass of the compound
LogP	Octanol–water partition coefficient
TPSA	Topological polar surface area
HBD	Number of hydrogen bond donors
HBA	Number of hydrogen bond acceptors
Atom Count	Total number of heavy atoms (non-hydrogen atoms)
Rotatable Bonds	Number of rotatable bonds (molecular flexibility)
MR	Molecular refractivity

3.2 Application of Drug-Likeness Filters

Table 2: Drug-likeness filters and their criteria

Filter	Descriptor	Threshold
Lipinski	Molecular Weight	≤ 500 Da
	LogP	≤ 5
	HBD	≤ 5
	HBA	≤ 10
Ghose	Molecular Weight	160-480 Da
	LogP	-0.4-5.6
	Molar Refractivity	40-130
	Atom Count	20-70
Veber	TPSA	≤ 140 Å^2
	Rotatable Bonds	≤ 10
Egan	LogP	≤ 5.88
	TPSA	≤ 131 Å^2

3.3 Visualisation

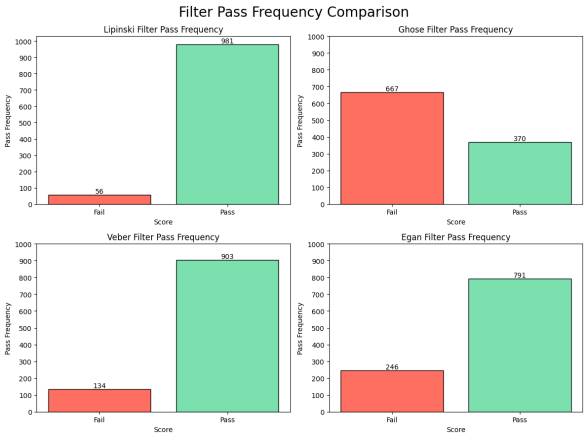

Figure 1: Filter-wise pass frequency comparison

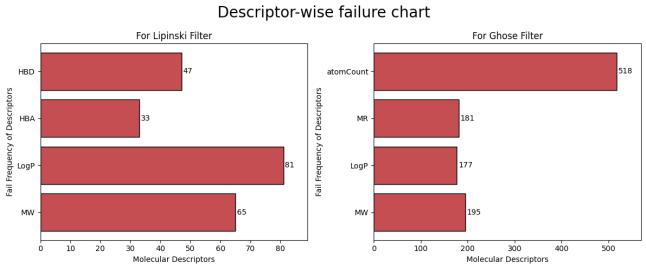

Figure 2: Lipinski and Ghose descriptor-wise fail frequency

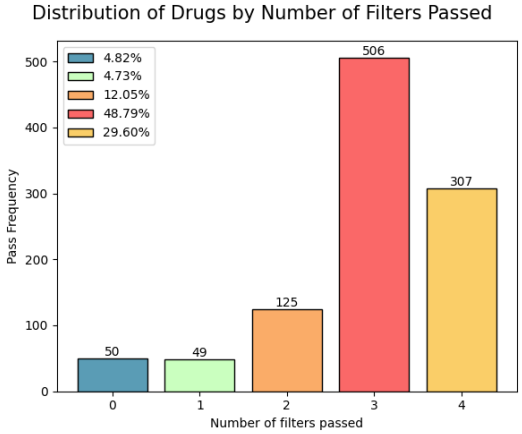

Figure 3: Distribution based on number of filters passed

4 Key Observations

1. From Fig.1, we observe that Lipinski and Veber filters exhibit similar and high permissiveness with over 900 molecules passing the filters. Egan filter imposes slightly stricter restrictions, with just under 800 molecules passing. On the contrary, Ghose filter has markedly strict filtering conditions, excluding a significantly larger fraction of compounds with 667 molecules failing. The rough agreement between the Lipinski, Veber and Egan filters suggest overlapping chemical spaces, aligning well with most of the approved drugs. A possible explanation for significant rejection rate for Ghose filter can be the extremely large or extremely small sizes of drug molecules. The specific descriptor-wise failure chart has been examined subsequently.

2. From Fig.2, it is evident that failure under Lipinski filter, lies in the LogP constraints, followed by the Molecular Weight conditions. Implying a stringent condition on drug liphophilicity. For Ghose filter, the dominant contributing factor responsible for the failure is the Atom Count condition, causing nearly 50% molecules from the dataset to fail, highlighting the stringent nature of this filter with respect to molecular size.

3. From Fig.3, it can be understood, that 49.08% of the dataset passes 3 out of 4 filters, followed by 29.60% passing all 4 filters and only 4.82% failing all four. This distribution indicates that the majority of FDA-approved drugs comply with most, but not necessarily all, established drug-likeness rules.

5 Conclusion

The overall study reflects agreement between the Lipinski, Veber and Egan filters and a significant disagreement and deviation for the Ghose filter. Further analysis shows the chances of failing all four filters is merely 4.82% whereas chances of passing at least 3 filters is 48.79%. Instances of failures can also stem from drugs targeting the central nervous system receptors, having significantly different properties than conventional oral drugs. This implies two major points -

- Most FDA-approved drugs comply with the 3 or more filters, thus emphasizing the viability of computationally analysed drug-likeness.

- However drugs tend to fail certain filters but pass certain others, making them filter dependent, rather than absolute.

The study reinforces, these physicochemical guidelines are only an early-screening tool and do not provide definite predictions for clinically successful drugs. Rigid adherence to these results can possibly result in rejection of potentially effective drug candidates.