# Do Zebrafish Obey Lipinski Rules?

Keith Long[†], Stephen J. Kostman[+], Christian Fernandez[†], James C. Burnett[‡], Donna M. Huryn[†]*

*[†]Department of Pharmaceutical Sciences, University of Pittsburgh, Pittsburgh, PA 15260; [+]Ohio Wesleyan University, Delaware, OH 43015; [‡]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260*



**Data Acquisition and Preparation**

Small molecules reported to be active in zebrafish assays were either extracted as SMILES or drawn in Chemdraw and converted to SMILES. Large peptides, and metal containing molecules were removed; salt ions were stripped from the SMILES.

Using ChemBioServer (Athanasiadis, E.; Cournia, Z.; Spyrou, G. ChemBioServer: a web-based pipeline for filtering, clustering and visualization of chemical compounds used in drug discovery. *Bioinformatics* **2012**, *28*, 3002-3003; http://chembioserver.vi-seem.eu), a list of 87 exemplars that represented the diversity of the zf 700 list was generated. To confirm the diversity of this set, an independent analysis using Discovery Studios was performed. Results from this analysis indicated that the average fingerprint distance = 0.90475 (with 1 = complete diversity and 0 = no diversity).

Using DrugBank 5.1.2 (https://www.drugbank.ca/releases/latest#structure), a 2D list of approved and experimental small molecule drugs was downloaded. The list was refined to include those with MW between 100 – 1500 containing at least one carbon atom, but no non-traditional elements except for B and Se. Salt ions were stripped for the analysis, but remain in the SMILES.

Data from Table 5, and SI in Reference 18 (Shultz, M.D. Two decades under the influence of the rule of five and the changing properties of approved oral drugs, *J. Med. Chem*. asap DOI: 10.1021/acs.jmedchem.8b00686) was used in Table 1 and for statistical analysis.

**Computational Methods**

Dassault Systèmes BIOVIA, Discovery Studio, 2018, San Diego: Dassault Systèmes, 2018 was used to calculate MW, AlogP, HBD, HBA, rotatable bonds, and molecular polar surface area.

Solubility, Caco-2 permeabilty, MDCK permeability and number of metabolites were predicted using the QikProp modules in Schrödinger Release 2016-1: Maestro, Schrödinger, LLC, New York, NY, 2018. Values for approximately 20 compounds out of 1395 within the DrugBank set could not be predicted.

MetStabOn (Podlewska, S.; Kafel, R. MetStabOn- online platform for metabolic stability prediction, *Int. J. Mol. Sci.* **2018**, 4, 1040-1057.) was used to predict human metabolic stability

**Statistical Analysis**

Excel from Microsoft Office 365 and R statistical software (version 3.2.3) was used to calculate means, standard deviations, 10$^{th}$ and 90$^{th}$ percentiles, minimums, and maximums for the zf 700, small molecule drugs, and the FDA approved oral NCEs datasets. Statistical significance between zf 700 values and the other two datasets was calculated by one-way ANOVA with Tukey's multiple comparison post hoc test or by Wilcoxon signed-rank test (for non-parametric data) within the entire and percentile datasets. All p values less than 0.05 were considered statistically significant (Values in xl files).

**Statistical Analysis of Percentiles**:

Table 2 Statistical Analysis:
Zebrafish 700 = zf
Small molecule Drugs = KD

| | QplogS | QPPCaco | QPPMDCK | metab |
|---|---|---|---|---|
| **90% zf** | -2.3128 | 4885.599 | 5229.392 | 7 |
| **90% KD** | -0.628 | 2163.85 | 1970.288 | 8 |
| **90% P value** | <2e-16 | <2e-16 | 8.50E-09 | 2.40E-07 |
| **10% zf** | -7.107 | 98.7296 | 56.5556 | 1 |
| **10% KD** | -6.5842 | 3.3938 | 2.2232 | 1 |
| **10% P value** | <2e-16 | <2e-16 | <2e-16 | <2e-16 |