

Accepted Manuscript

Title: Redundancy in two major compound databases

Authors: Dimitar Yonchev, Dilyana Dimova, Dagmar Stumpfe, Martin Vogt, Jürgen Bajorath

PII: S1359-6446(18)30027-8
DOI: <https://doi.org/10.1016/j.drudis.2018.03.005>
Reference: DRUDIS 2204

To appear in:



Please cite this article as: Yonchev, Dimitar, Dimova, Dilyana, Stumpfe, Dagmar, Vogt, Martin, Bajorath, Jürgen, Redundancy in two major compound databases. Drug Discovery Today <https://doi.org/10.1016/j.drudis.2018.03.005>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Redundancy in two major compound databases

Dimitar Yonchev, Dilyana Dimova, Dagmar Stumpfe, Martin Vogt and Jürgen Bajorath*

Department of Life Science Informatics, B-IT, LIMES Program Unit Chemical Biology and Medicinal Chemistry, Rheinische Friedrich-Wilhelms-Universität, D-53113 Bonn, Germany

**Corresponding author. Bajorath, J. (bajorath@bit.uni-bonn.de).*

Highlights:

- Public compound databases are of particular relevance for drug discovery research
- Crosstalk between databases is desirable but also leads to redundancy
- ChEMBL and PubChem contain large numbers of shared compounds and activity records
- More than half of the 14.7 million activities in ChEMBL originate from PubChem
- More than 99% of the 1.25 million assays in PubChem are imported from ChEMBL
- Data exchange requires a high level of transparency

Download English Version:

<https://daneshyari.com/en/article/8409569>

Download Persian Version:

<https://daneshyari.com/article/8409569>

[Daneshyari.com](https://daneshyari.com)