Full Paper

www.molinf.com

molecular informatics

DOI: 10.1002/minf.201700029

Comprehensive Network Map of ADME-Tox Databases

Baptiste Canault,[a] Stéphane Bourg,[a] Philippe Vayer,[b] and Pascal Bonnet*[a]

Abstract: In the last decade, many statistical-based approaches have been developed to improve poor pharmacokinetics (PK) and to reduce toxicity of lead compounds, which are one of the main causes of high failure rate in drug development. Predictive QSAR models are not always very efficient due to the low number of available biological data and the differences in the experimental protocols. Fortunately, the number of available databases continues to grow every year. However, it remains a challenge to determine the source and the quality of the original data. The main goal is to identify the relevant databases required to generate the most robust predictive models. In this study, an interactive network of databases was proposed to easily find online data sources related to ADME-Tox parameters data. In this map, relevant information regarding scope of application, data availability and data redundancy can be obtained for each data source. To illustrate the usage of data mining from the network, a dataset on plasma protein binding is selected based on various sources such as DrugBank, PubChem and ChEMBL databases. A total of 2,606 unique molecules with experimental values of PPB were extracted and can constitute a consistent dataset for QSAR modeling.

Keywords: ADME-Tox · Network · Database · PPB

1 Introduction

The pharmaceutical industry is facing important challenges through every stage of the drug discovery process. Nowadays, computer-assisted drug design plays a key role in decision-making from the start of R&D projects until selection of clinical candidate.[1] A drug needs to be safe, effective, and have appropriate pharmacokinetic properties (PK). Recent studies indicate that inappropriate PK properties for drug candidates are one of the main causes for late-stage failures especially in phase I clinical trials.[2] The prediction of absorption, distribution, metabolism, excretion (ADME) and toxicity (Tox) properties is a research field very investigated early in lead optimization. Accordingly, prediction models were developed for human PK to predict a wide range of chemical structures and to guide medicinal chemists in both design and optimization of improved chemical leads such as quantitative structure-activity/property relationship (QSAR or QSPR).[3,4] These computational models are based on experimental data and seek to cover a widest possible chemical space with a reliable prediction. The quantity and the quality of available data used to train models play a crucial role in the coverage of chemical space and performance of the model.[5,6] Nevertheless, some homogeneous *in vitro* or *in vivo* ADME-Tox data are subject to commercial licenses and/or usually concern private chemical collections.[7] However, while some ADME-Tox data have been already published they are still sparse throughout a number of databases, which creates difficulties in constituting large datasets in order to perform robust and reliable computational models.[8] Today, a multitude of online life sciences databases have emerged and could also contribute to the elaboration of larger datasets than those already published. Nevertheless, numerous issues are raised when scientists want to collect relevant ADME-Tox data from online databases.[9]

Most databases integrate a wide variety of numerical data and cannot be easily assigned to a specific area of research. Although a minority of databases is domain-specific, many of them provide some data applicable to multidisciplinary domains of research. ADME-Tox information represent a very small part of data commonly found on the web, it is therefore not surprising that databases are not clearly designated as providing data related to PK properties. As an example, the Database Commons Catalog (http://databasecommons.org) and the Nucleic Acids Research online Database Collection[10] do not take into account ADME-Tox category to classify databases. Consequently, it remains more difficult to locate easily databases for a specific application related to pharmacokinetics.

All databases accessible online are not necessarily available. Although many resources are freely available, some databases are commercial or only accessible online without any possibility to download the data. The data availability can be a serious impediment to collect data, standardize data and develop predictive models.

[a] *B. Canault, S. Bourg, P. Bonnet*
Institut de Chimie Organique et Analytique (ICOA), Université d'Orléans et CNRS, UMR7311, BP 6759, 45067 Orléans, France
E-mail: pascal.bonnet@univ-orleans.fr
[b] *P. Vayer*
Technologie Servier
25-27 rue Eugène Vignat, BP 11749, 45007 Orléans cedex 1 (France)

Supporting information for this article is available on the WWW under https://doi.org/10.1002/minf.201700029

Information on molecular structures and/or experimental data are present in many databases. Therefore, the risks of data redundancy and errors propagation from supplier to client databases are growing.[11] Understanding the relationships between online resources could be crucial in order to save time and to locate the most useful database for specific needs. The need to have such global network of interconnected life sciences-related databases and more precisely a global ADME-Tox network has been previously discussed,[12] but it has never been implemented and proposed to the scientific community.

While some databases contain highly curated data, some only provide raw data without careful analysis and critical assessment. Consequently, data quality can strongly vary between databases created from different sources. Nevertheless, information on data curation and data quality is not always clearly defined and it remains difficult to determine whether or not collected data may contain erroneous measurements.

In the present work, a comprehensive network map of ADME-Tox data sources was proposed to offer a global understanding of the relationships between them and to easily retrieve data for a specific need. Particular attention was given to the description of data content and availability. A network analysis was proposed to determine the level of connections between databases and characterize their role of supplier or client in the proposed network. Finally, in order to demonstrate the applicability of the network, we proposed a case study with the selection of relevant ADME-Tox databases in order to constitute a large dataset including plasma protein binding (PPB) property. While not performed in the current study, this dataset could be used to build QSAR models for ADME-Tox modeling.

2 Materials and Methods

2.1 Creation of ADME-Tox Network

The coherent map was performed with the R package visNetwork.[13] This package is intuitive and fast for the visualization of dynamic network composed of several thousand objects. To build ADME-Tox network, this tool requires sets of nodes (*N*) and edges (*E*) representing all the databases and their connections, respectively.

The set *N* was generated with a manual search extracted from VLS3D (http://www.vls3d.com) and Click2Drug (https://www.click2drug.org), which are two comprehensive lists of 150 databases or tools related to ADME-Tox applications. We described each database by a validated URL, by a field corresponding to data availability (yes or no) and by their corresponding reference (DOI or PMID). A classification based on the Nucleic Acids Research online Database Collection[10] was added to the *N* set, as category or sub-category nodes, and represents the scope of database according to their content. This classification was completed with "ADME-Tox" category and five sub-categories "Absorption", "Distribution", "Metabolism", "Elimination" and "Toxicity".

Information about origin of data sources was manually extracted from articles of published databases, or from database websites when available. This information constitutes the edges between nodes (set *E)* containing links of reference source (client or supplier) between databases. Moreover, we added connections between databases and category or sub-category nodes that help to characterize multidisciplinary content of data resource directly on the network map. In some cases, the analysis of the origin of the data has allowed us to identify new databases (not included in VLS3D or Click2Drug lists), which were then added to the set of nodes (set *E*).

2.2 Network Analysis

After the construction of the network, we removed the databases, which are not connected to any categories or sub-categories. They correspond to unexploitable resources in the network because of their single edge (network boundary) and consequently these databases are not relevant to perform analysis. A sub-network was obtained, which contains only databases with an exhaustive description of their data content. From this reduced network, we took into account all the links present between databases and the orientation of the edges in order to characterize the role, as supplier or client, of each connected ADME-Tox data source. The edges between databases and categories were not counted since it does not provide any relevant information and would overestimate the number of links.

For each database, we calculated the number of nodes corresponding to the sum of the number of suppliers and clients providing or requesting information respectively. Finally, a database was considered as "Supplier", if the number of suppliers is higher than the number of clients, otherwise it was defined as "Client".

2.3 Dataset Preparation

Chemical structures can be drawn according to different rules, which results in different drawings of structures from diverse databases corresponding to the same compound.[14] To address this issue, we have standardized all chemical structures extracted from selected databases with the following protocol: mixture, salts and solvent were removed, charges were neutralized and molecular structures were cleaned using ChemAxon standardization tools.[15] The standardization was complete by the protonation of the structures at pH 7.4. Standard InChI and standard InChIKey were calculated from standardized chemical structures. When multiple values of a property were present in a dataset, we selected the maximal value.

3 Results

3.1 Presentation of ADME-Tox Network

The complete network, presented in Figure 1.A, contains 11 categories, 29 sub-categories and 373 databases connected by 865 edges. Different features were applied to facilitate understanding of the network (Figure 1.B). Grey squares represent categories and sub-categories from the online Database Collection while orange squares and dots represent the newly added ADME-tox data (Figure 1.B). The links are oriented depending on the original source and represent the flow of diverse data between "client" and "supplier" databases (Figure 1.C). The network was implemented in a dynamic web page containing some widgets to help the user in the selection of subsets of nodes using database or category name as presented in supporting information.

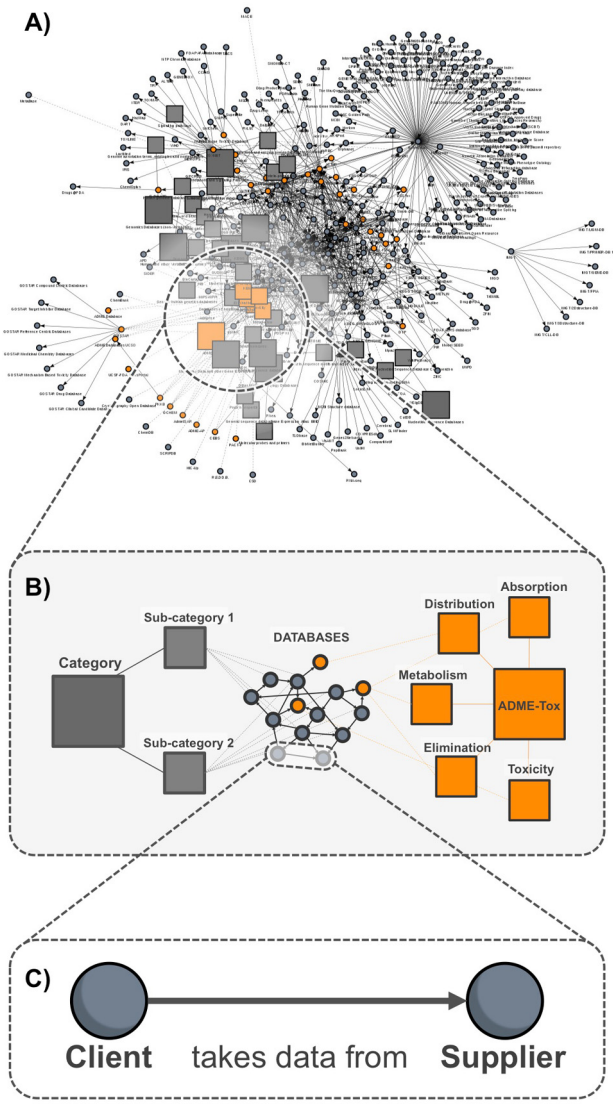

Figure 1. ADME-Tox network. (A) Complete network representation with defined connections between databases manually extracted from published articles. (B) Description of ADME-Tox network features. The "square" and "dot" differentiate categories and databases respectively; node size discerns categories from sub-categories, as well as orange color distinguishes ADME-Tox databases from other nodes. (C) Orientation of edges in the ADME-Tox network.

3.2 Analysis of ADME-Tox Databases Connections

The network analysis was performed using sub-network resulting from data preparation. A total of 267 databases, which are not connected to any categories or sub-categories, were removed. The reduced network and all information included in the databases related to ADME-Tox properties are provided in supporting information.

Determine the most active databases in term of data exchanges and highlight databases according to their global roles are two key factors to locate the most useful data sources. In this study, the number of connections and orientations of the links are used to evaluate the global role of each ADME-Tox database. The number of "Clients" and "Suppliers" of each database and their respective connections are shown in Figure 2. The majority of the most connected databases are those proposing downloadable datasets. Accordingly, three groups are now identified in

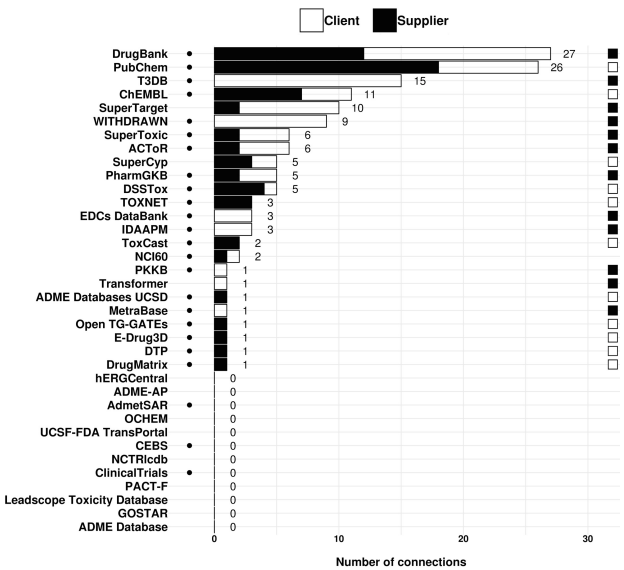

Figure 2. Client and Supplier databases used in the network. Bar chart corresponds to the total number of connections per database, where white and black colors represent the number of suppliers and clients, respectively. The squares represent the main groups ("Supplier", "Client") of the databases in the network. The black filled circles indicate whether a database provides a downloadable dataset. The empty circles represent undownloadable data from the database.

the network such as "Supplier", "Client" and "Undefined" groups.

3.3 Case Study: Preparation of Dataset of Plasma Protein Binding

Drugs present in the blood circulation are in reversible association with plasma proteins. The ratio unbound/(unbound + bound) characterizes the fraction of free drug available to induce the therapeutic effect. The unbound fraction of the drug is also the portion that may be metabolized and/or eliminated. As a reservoir, the bound fraction is released to maintain the equilibrium and therefore is related to the metabolic half-life of a drug in the body. Therefore, Plasma Protein Binding (PPB) is a key property of drug distribution, which can also impact pharmacological properties of the drugs.

Many questions necessarily arise when searching data for a specific ADME-Tox application, e.g., where are the databases that contain data available to my needs? Are the data easily accessible? Are the data redundant and/or reliable? In this section, we will show how the network map of ADME-Tox databases can help to locate and collect several data to constitute a large dataset related to PPB.

We used the reduced network to identify 16 potential sources of PPB data in the "Distribution" sub-category. After careful analysis, only 7 databases contained PPB data. AdmetSAR,[16] PharmGKB[17] and E-Drug3D[18] databases were excluded because their PPB data is not accessible as a downloadable dataset. ClinicalTrials[19] database was discarded due to the low number of accessible assays with known chemical structures. Only data from ChEMBL,[20] PubChem[21] and DrugBank[22] databases were extracted and prepared. A total of 3,436 molecules with PPB values were retrieved.

In order to control the chemical structures redundancy, we used the Venn diagram shown in Figure 3.A, which represents the overlap between datasets using the standard InChIKey. Only 37 molecules are common to all datasets and ChEMBL covers larger parts of PubChem with 754 overlapped molecules. Furthermore, DrugBank has 756 unique molecules not present in the other two databases and 76 molecules in common. To see how DrugBank is different from PubChem and ChEMBL, we compared PPB distribution of each dataset as depicted in Figure 3.B. Whereas ChEMBL, PubChem and DrugBank are left-skewed for the PPB distribution (medians around 95%), DrugBank covers a wider range of PPB values since 50% of the values are between 50 and 98% of PPB. Nevertheless, the representation of the 37 common molecules based on their identical InChIKey shows that DrugBank values are different from the two other databases on the boxplots (white circle).

The data reliability is another important issue to provide quality data for the development of reliable models. Figure 3.C depicts the PPB difference between each dataset using 28 bins. Among the 754 molecules in common between PubChem and ChEMBL, PPB values are approximately 87% identical and 10% with a difference between 1 and 10% of PPB and only 3% are different with a range of difference of binned PPB (△PPB) higher than 10%. According to the results of the overlap and PPB difference between datasets, ChEMBL and PubChem are roughly equivalent in term of PPB measurements whereas DrugBank is significantly different from ChEMBL and PubChem with only 37% and 42% of identical PPB values respectively.

4 Discussion

Nowadays, scientists are faced with the increase of dataflow widespread in various databases. It becomes difficult to locate relevant ADME-Tox data across the multitude of available life science databases. To address this issue, we propose a comprehensive and interactive network using 373 connected databases (Figure 1.A). On this map, nodes of categories and sub-categories were added to help users to easily identify relevant databases. In this regard, information about data availability was added for each database and this information could save time during data collection in a research project.

We focused our work on a sub-network to allow a new understanding of connections between commonly used databases and attempts to offer an overview of the available PK data resources. As we noted above, the most used databases propose downloadable datasets and three groups of databases are observed in Figure 2. The "Client" group is composed of DrugBank,[22] T3DB,[23] SuperTarget,[24] WITHDRAWN,[25] ACToR,[26] PharmGKB,[17] SuperToxic,[27] EDCs DataBank,[28] IDAAPM,[29] PKKB,[30] Transformer,[31] and finally MetraBase.[32] Most of them are public databases containing multidisciplinary accessible data from diverse sources. Data contained in these databases are generally curated as far as possible, which make them suitable to build a dataset for predictive modeling. The "Supplier" group contained a majority of databases that are supported by governmental agencies, such as PubChem,[21] ChEMBL,[20] DSSTox,[33] TOXNET,[34] ToxCast,[35] Open TG-GATEs,[36] DTP,[37] or even DrugMatrix.[38] Most of them contains *in vitro* and/or *in vivo* assays, as well as information about experimental protocols, which is important to create homogeneous datasets.[39] The third group contained databases with undefined role in the network. This is explained by either a perfect balance in the flow of data exchanges of "Supplier" and "Client" like NCI-60 from DTP or a total absence of connections. Concerning this last point, some databases are public and downloadable but are surprisingly not used as data provider for other databases, such as AdmetSAR,[16] CEBS[40] and ClinicalTrials.[19] Others are public but with an online restricted access like OCHEM,[41] UCSF-FDA TransPortal,[42] hERGCentral[43] or ADME-AP,[44] and finally some are private with restricted data access like GOSTAR,[45] Leadscope,[46] ADME Database[47]

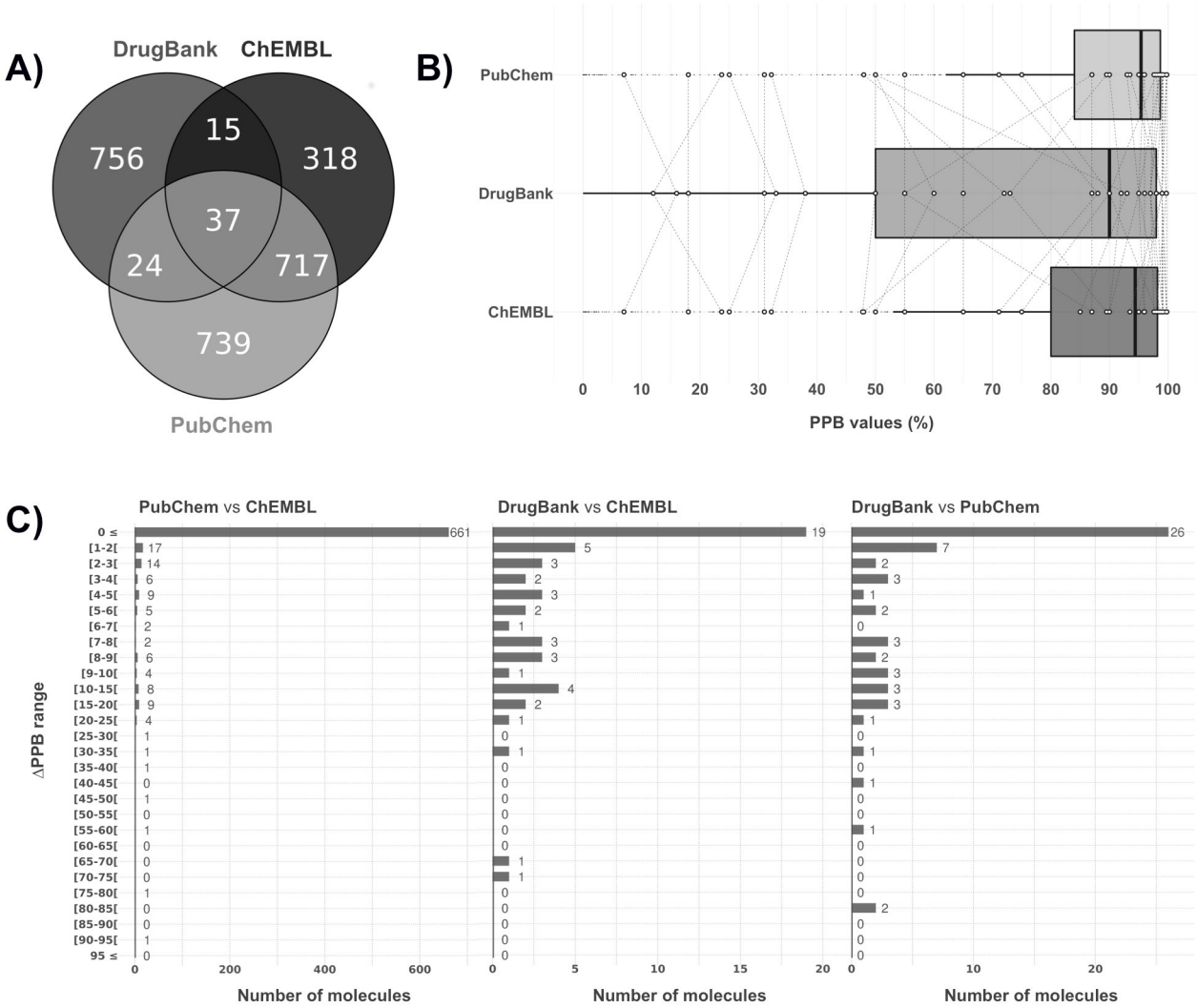

Figure 3. Analysis of PPB datasets extracted from ChEMBL, DrugBank and PubChem. (A) Venn diagram comparing molecules in each dataset. (B) Box plots of PPB distribution with representation of the 37 common molecules (white circle) in the 3 databases. (C) Difference (△PPB) of binned PPB values between ChEMBL, DrugBank and PubChem.

or PACT-F.[48] It should be noted that the most important commercial databases like Cloe Knowledge[49] and Pharmapendium[50] are not present in the network because they were not present in VLS3D and Click2Drug lists. Consequently, this network represents the most common database in the field of life sciences and gives us the possibility to explore all the gathered data in a unique interactive map.

To know if the present work was useful to find and collect relevant data, we proposed a case study involving the collection of a PPB dataset. The datasets were extracted from ChEMBL, PubChem and DrugBank that have been previously identified as major supplier databases (PubChem and ChEMBL) or major client database (DrugBank). While the number of common molecules between DrugBank and the other two databases is small (Figure 3.A), this case study shows that ChEMBL and PubChem contain a large number of redundant data, despite few different PPB values (Figure 3.C). This observation is in accordance with the known complementarity information between ChEMBL and PubChem,[51] which is represented on the network by a balanced edge (reduced network Supporting Information). Therefore, the network orientation of links between these two databases provides information on the data redundancy. Regarding to data reliability, DrugBank provides different PPB values for 76 common structures with the other databases and the majority of △PPB are ranging between 0 and 20%. By combining various databases into a unique dataset, care should be taken on the pretreatment of the data, and a careful cleaning process is required.

Moreover, 756 molecules remain exclusive to DrugBank. The analysis of the data shows that DrugBank provides complementary data to PubChem and ChEMBL. Thus, a total of 2,606 unique molecules containing experimental values, extracted from these three databases, could be used to predict PPB property whereas the majority of QSAR models already published are based on datasets with a number of molecules below 1,000 compounds.[52–59]

5 Conclusions

In this study, a comprehensive map of ADME-Tox databases extracted from a network of life science databases is proposed. A case study is presented using data extraction of PPB values which resulted in 2,606 molecules with corresponding experimental values. There are several advantages in using this network. First, the present work combines and describes the most commonly used databases in the field of life science and especially in pharmacokinetics. Furthermore, the interactive network map combines all information necessary to understand connections between databases and can be used to select the available resources for a specific application. Finally, the network can be useful in analyzing data redundancy by visualizing the dataflow between data sources. This comprehensive network map is a tool that can find useful application in drug discovery projects and especially in the ADME-Tox research field.

Conflict of Interest

None declared.

Acknowledgements

This work was also supported by Servier. Authors whish to thank the Région Centre Val de Loire for financial supports.

References

[1] G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, *Pharmacol. Rev.* **2013**, *66*, 334–395.
[2] M. J. Waring, J. Arrowsmith, A. R. Leach, P. D. Leeson, S. Mandrell, R. M. Owen, G. Pairaudeau, W. D. Pennie, S. D. Pickett, J. Wang, et al., *Nat. Rev. Drug Discov.* **2015**, *14*, 475–486.
[3] J. Hodgson, *Nat. Biotechnol.* **2001**, *19*, 722–726.
[4] A. Boobis, U. Gundert-Remy, P. Kremers, P. Macheras, O. Pelkonen, *Eur. J. Pharm. Sci. Off. J. Eur. Fed. Pharm. Sci.* **2002**, *17*, 183–193.
[5] J. Kirchmair, A. H. Göller, D. Lang, J. Kunze, B. Testa, I. D. Wilson, R. C. Glen, G. Schneider, *Nat. Rev. Drug Discov.* **2015**, *14*, 387–404.
[6] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, et al., *J. Med. Chem.* **2014**, *57*, 4977–5010.
[7] I. V. Tetko, O. Engkvist, U. Koch, J.-L. Reymond, H. Chen, *Mol. Inform.* **2016**, *35*, 615–621.
[8] J. Bajorath, *F1000Research* **2015**, *4*, DOI 10.12688/f1000research.6653.1.
[9] R. Guha, D.-T. Nguyen, N. Southall, A. Jadhav, *Curr. Protoc. Chem. Biol.* **2012**, *4*, 193–209.
[10] D. J. Rigden, X. M. Fernández-Suárez, M. Y. Galperin, *Nucleic Acids Res.* **2016**, *44*, D1-6.
[11] S. Philippi, J. Köhler, *Nat. Rev. Genet.* **2006**, *7*, 482–488.
[12] H. Van de Waterbeemd, M. De Groot, *SAR QSAR Environ. Res.* **2002**, *13*, 391–401.
[13] A. B. V. (vis js library in htmlwidgets/lib, http://visjs.org, http://www.almende.com/home), B. T. (R interface), *visNetwork: Network Visualization Using "Vis.js" Library*, **2016**.
[14] A. Hersey, J. Chambers, L. Bellis, A. Patrícia Bento, A. Gaulton, J. P. Overington, *Drug Discov. Today Technol.* **2015**, *14*, 17–24.
[15] JChem 15.2.9, ChemAxon, **2015**, http://www.chemaxon.com.
[16] F. Cheng, W. Li, Y. Zhou, J. Shen, Z. Wu, G. Liu, P. W. Lee, Y. Tang, *J. Chem. Inf. Model.* **2012**, *52*, 3099–3105.
[17] M. Hewett, D. E. Oliver, D. L. Rubin, K. L. Easton, J. M. Stuart, R. B. Altman, T. E. Klein, *Nucleic Acids Res.* **2002**, *30*, 163–165.
[18] E. Pihan, L. Colliandre, J.-F. Guichou, D. Douguet, *Bioinforma. Oxf. Engl.* **2012**, *28*, 1540–1541.
[19] NLM, NIH, "ClinicalTrials," can be found under https://clinicaltrials.gov, **2017**.
[20] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, et al., *Nucleic Acids Res.* **2012**, *40*, D1100-1107.
[21] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, et al., *Nucleic Acids Res.* **2016**, *44*, D1202-1213.
[22] D. S. Wishart, C. Knox, A. C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, Z. Chang, J. Woolsey, *Nucleic Acids Res.* **2006**, *34*, D668-672.
[23] E. Lim, A. Pon, Y. Djoumbou, C. Knox, S. Shrivastava, A. C. Guo, V. Neveu, D. S. Wishart, *Nucleic Acids Res.* **2010**, *38*, D781-786.
[24] N. Hecker, J. Ahmed, J. von Eichborn, M. Dunkel, K. Macha, A. Eckert, M. K. Gilson, P. E. Bourne, R. Preissner, *Nucleic Acids Res.* **2012**, *40*, D1113-1117.
[25] V. B. Siramshetty, J. Nickel, C. Omieczynski, B.-O. Gohlke, M. N. Drwal, R. Preissner, *Nucleic Acids Res.* **2016**, *44*, D1080-1086.
[26] R. Judson, A. Richard, D. Dix, K. Houck, F. Elloumi, M. Martin, T. Cathey, T. R. Transue, R. Spencer, M. Wolf, *Toxicol. Appl. Pharmacol.* **2008**, *233*, 7–13.
[27] U. Schmidt, S. Struck, B. Gruening, J. Hossbach, I. S. Jaeger, R. Parol, U. Lindequist, E. Teuscher, R. Preissner, *Nucleic Acids Res.* **2009**, *37*, D295-299.
[28] D. Montes-Grajales, J. Olivero-Verbel, *Toxicology* **2015**, *327*, 87–94.
[29] A. Legehar, H. Xhaard, L. Ghemtio, *J. Cheminformatics* **2016**, *8*, 33.
[30] D. Cao, J. Wang, R. Zhou, Y. Li, H. Yu, T. Hou, *J. Chem. Inf. Model.* **2012**, *52*, 1132–1137.
[31] M. F. Hoffmann, S. C. Preissner, J. Nickel, M. Dunkel, R. Preissner, S. Preissner, *Nucleic Acids Res.* **2014**, *42*, D1113-1117.
[32] L. Mak, D. Marcus, A. Howlett, G. Yarova, G. Duchateau, W. Klaffke, A. Bender, R. C. Glen, *J. Cheminformatics* **2015**, *7*, 31.
[33] US EPA, "DSSTox," can be found under https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database, **2017**.

[34] G. C. Fonger, D. Stroup, P. L. Thomas, P. Wexler, *Toxicol. Ind. Health* **2000**, *16*, 4–6.
[35] US EPA, "ToxCast," can be found under https://www.epa.gov/chemical-research/toxicity-forecasting, **2017**.
[36] Y. Igarashi, N. Nakatsu, T. Yamashita, A. Ono, Y. Ohno, T. Urushidani, H. Yamada, *Nucleic Acids Res.* **2015**, *43*, D921–927.
[37] NIH, "Developmental Therapeutics Program (DTP)," can be found under https://dtp.cancer.gov, **2017**.
[38] B. Ganter, S. Tugendreich, C. I. Pearson, E. Ayanoglu, S. Baumhueter, K. A. Bostian, L. Brady, L. J. Browne, J. T. Calvin, G.-J. Day, et al., *J. Biotechnol.* **2005**, *119*, 219–244.
[39] X. Fu, A. Wojak, D. Neagu, M. Ridley, K. Travis, *J. Cheminformatics* **2011**, *3*, 24.
[40] NIH, "Chemical Effects in Biological Systems (CEBS)," can be found under http://tools.niehs.nih.gov/cebs3/ui/, **n.d.**
[41] I. Sushko, S. Novotarskyi, R. Körner, A. K. Pandey, M. Rupp, W. Teetz, S. Brandmaier, A. Abdelaziz, V. V. Prokopenko, V. Y. Tanchuk, et al., *J. Comput. Aided Mol. Des.* **2011**, *25*, 533–554.
[42] K. M. Morrissey, C. C. Wen, S. J. Johns, L. Zhang, S.-M. Huang, K. M. Giacomini, *Clin. Pharmacol. Ther.* **2012**, *92*, 545–546.
[43] F. Du, H. Yu, B. Zou, J. Babcock, S. Long, M. Li, *Assay Drug Dev. Technol.* **2011**, *9*, 580–588.
[44] L. Z. Sun, Z. L. Ji, X. Chen, J. F. Wang, Y. Z. Chen, *Bioinforma. Oxf. Engl.* **2002**, *18*, 1699–1700.
[45] S. A. R. P. Jagarlapudi, K. V. R. Kishan, *Methods Mol. Biol. Clifton NJ* **2009**, *575*, 159–172.
[46] C. Yang, R. D. Benz, M. A. Cheeseman, *Curr. Opin. Drug Discov. Devel.* **2006**, *9*, 124–133.
[47] T. Fujitsu Kyushu Systems, "ADME Database," can be found under http://www.fujitsu.com/jp/group/kyushu/en/solutions/industry/lifescience/admedatabase, **2017**.
[48] PharmaInformatic, "PACT-F," can be found under http://www.pharmainformatic.com/html/pact-f.html, **2017**.
[49] Cyprotex, "CloeKnowledge," can be found under http://www.cloegateway.com, **2017**.
[50] Elsevier, "PharmaPendium," can be found under https://www.elsevier.com/solutions/pharmapendium-clinical-data, **2017**.
[51] A. P. Bento, A. Gaulton, A. Hersey, L. J. Bellis, J. Chambers, M. Davies, F. A. Krüger, Y. Light, L. Mak, S. McGlinchey, et al., *Nucleic Acids Res.* **2014**, *42*, D1083–1090.
[52] X.-W. Zhu, A. Sedykh, H. Zhu, S.-S. Liu, A. Tropsha, *Pharm. Res.* **2013**, *30*, 1790–1798.
[53] E. M. del Amo, L. Ghemtio, H. Xhaard, M. Yliperttula, A. Urtti, H. Kidron, *PloS One* **2013**, *8*, e74758.
[54] M. Lobell, V. Sivarajah, *Mol. Divers.* **2003**, *7*, 69–87.
[55] J. Wang, G. Krudy, X.-Q. Xie, C. Wu, G. Holland, *J. Chem. Inf. Model.* **2006**, *46*, 2674–2683.
[56] S. Weaver, M. P. Gleeson, *J. Mol. Graph. Model.* **2008**, *26*, 1315–1326.
[57] H. Li, Z. Chen, X. Xu, X. Sui, T. Guo, W. Liu, J. Zhang, *Biopharm. Drug Dispos.* **2011**, *32*, 333–342.
[58] K. Yamazaki, M. Kanaoka, *J. Pharm. Sci.* **2004**, *93*, 1480–1494.
[59] J. R. Votano, M. Parham, L. M. Hall, L. H. Hall, L. B. Kier, S. Oloff, A. Tropsha, *J. Med. Chem.* **2006**, *49*, 7169–7181.

Received: February 10, 2017
Accepted: June 14, 2017
Published online on June 29, 2017

