analytical chemistry

Technical Note

Cite This: Anal. Chem. 2019, 91, 4312−4316

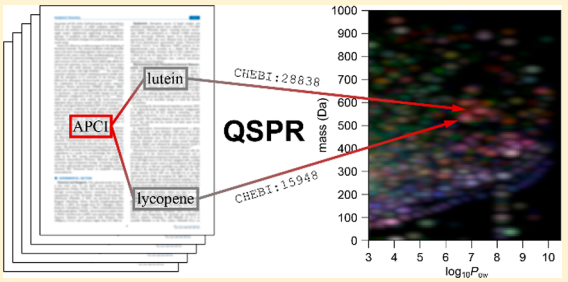
pubs.acs.org/ac

Visual and Semantic Enrichment of Analytical Chemistry Literature Searches by Combining Text Mining and Computational Chemistry

Magnus Palmblad*

Center for Proteomics and Metabolomics, Leiden University Medical Center, Postzone S3-P, Postbus 9600, 2300 RC Leiden, The Netherlands

ABSTRACT: The open-access scientific literature contains a wealth of information for meaningful text mining. However, this information is not always easy to retrieve. This technical note addresses the problem by a new flexible method combining in a single workflow existing resources for literature searches, text mining, and large-scale prediction of physicochemical and biological properties. The results are visualized as virtual mass spectra, chromatograms, or images in styles new to text mining but familiar to analytical chemistry. The method is demonstrated on comparisons of analytical-chemistry techniques and semantically enriched searches for proteins and their activities, but it may also be of general utility in experimental design, drug discovery, chemical syntheses, business intelligence, and historical studies. The method is realized in shareable scientific workflows using only freely available data, services, and software that scale to millions of publications and named chemical entities in the literature.



The scientific literature provides an abundance of information in the public domain for text mining and machine learning, including PubMed with 29.1 million titles and abstracts and Europe PMC[1] with 2.25 million full-text articles that are open access or have CC-BY, CC-BY-NC, or CC0 licenses. Millions of named entities such as diseases, genes, proteins, and metabolites are already text-mined and annotated in the SciLite[2] platform. These annotations can be accessed programmatically from Europe PMC via the recently introduced Annotations API, which is similar to the web services used to access bibliographic information.[1] These services can be combined and the searches orchestrated in scientific workflow systems such as Taverna[3−6] or KNIME,[7] which also ensures reproducibility and enables open sharing of literature analyses.

Metabolites, drugs, and other small molecules are annotated using the Chemical Entities of Biological Interest (ChEBI) ontology[8] containing 99 413 entities (release 169), along with synonyms, CAS registry numbers, elemental formulas, masses, and chemical structures in SMILES[9] and InChI[10] formats. These structures are used to derive a number of molecular descriptors. From these descriptors and available experimental data, machine learning is used to predict important physicochemical and biological properties, such as aqueous solubility,[11−13] melting point,[11,14] vapor pressure,[15,16] bioavailability,[17,18] developmental toxicity,[19,20] and receptor binding.[21]

This note describes how text mining for chemical entities can be combined with prediction of their physicochemical properties to guide the selection of analytical methods, extract information on proteins and their ligands, or study the history of a particular subfield of chemistry in semantically enriched literature searches.

EXPERIMENTAL SECTION

The method presented here is a combination of three web-based services made interoperable through scripts embedded in a workflow executed on the user's side. The starting point is one or a set of literature queries. Each query is used to generate one API call to the Europe PMC searchPublications RESTful service. The searches can be restricted to previously tagged[22] sections, such as the methods sections, or unrestricted. Each call returns a list of PubMed identifiers with metadata such as whether annotations from text mining are available. The annotated chemical compounds and classes in publications for which text mining results are available are retrieved using the recently launched getAnnotations service (Figure 1). After these searches, molecular information such as mass, elemental composition, and structures (as SMILES or InChIs) can be looked up in the ontology from the returned molecular (ChEBI) identifiers. The returned SMILES can then be used to predict a number of physicochemical and biological properties on a computational chemistry server such as OCHEM.[23] Polarities were predicted using ALogPS 3.0 on OCHEM as the 1-octanol−water-partition coefficient (log P). Estrogen-receptor and p53-signaling agonists were predicted using the qualitative consensus estrogen-receptor-α- and p53-signaling-agonists OCHEM models[21] (model IDs 518 and 522, respectively) developed to answer the Tox21 challenge.[24] The 4216 ChEBI compounds that had already been used to train the models for ER and p53-signaling agonists were removed

Received: December 18, 2018
Accepted: March 5, 2019
Published: March 5, 2019

DOI: 10.1021/acs.analchem.8b05818
Anal. Chem. 2019, 91, 4312−4316

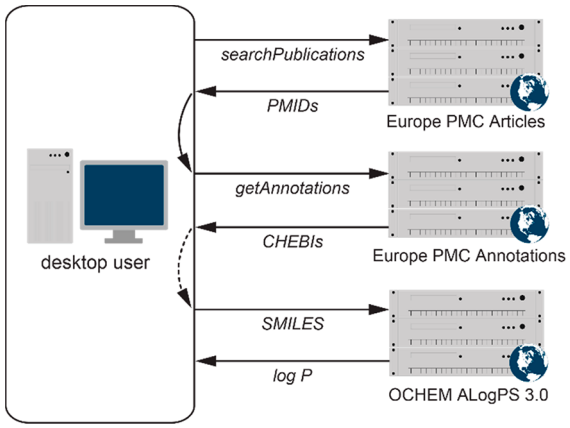


Figure 1. Schematic of analysis. In principle, all analysis in this note can be done by connecting three web services or applications: searchPublications and getAnnotations from Europe PMC and the OCHEM ALogPS prediction model. The service calls can be orchestrated by a single scientific workflow in workflow managers such as Taverna or KNIME.

from further analysis. The ChEBI ontology includes 5104 chemical classes with wildcards in their SMILES, individual elements, isotopes, and even elementary particles. As physicochemical properties cannot be meaningfully measured or predicted for such, these ChEBI entities were also excluded, even though they are annotated in many publications. To accelerate the workflow, these properties were precomputed for the 99 413 entities in ChEBI and stored locally in a lookup table. The ChEBI ontology is updated monthly, whereas Europe PMC results can change on a daily basis with new publications and text-mined items. The molecular masses were extracted from the ChEBI OBO file and added to the lookup table.

To compare multiple queries, the results were normalized against the total number of annotations for each comparand. Visualizations were generated in R. For density plots, a 2D Gaussian blur was applied with standard deviations 0.125 log *P* units (two bins) and 6.25 Da (one bin) to represent the inherent uncertainties in log *P* predictions as well as the difference between calculated monoisotopic masses and measured masses due to the presence of isotopes, charge carriers, adducts, or neutral losses. The custom computer code used in this work is embedded in a KNIME workflow available on GitHub (https://github.com/magnuspalmblad/EuropePMC2ChEBI). The R functions used to visualize the results are also included in the repository.

To demonstrate the methodology, Europe PMC was queried for "electrospray ionization" (ESI), "atmospheric pressure chemical ionization" (APCI), "electron impact", "and gas chromatography" (GC). Furthermore, tagged methods sections were searched for "liquid chromatography–mass spectrometry" (LC-MS). Europe PMC was also searched for annotations of the estrogen receptor (UniProt P03372) and the cellular tumor antigen p53 (UniProt P04637).

RESULTS

The results from a single literature search can be visualized as a distribution of one or more derived properties of the chemical compounds retrieved from the text-mined publications. Figure 2 illustrates this as a familiar "mass spectrum" for all compounds annotated in publications mentioning LC-MS in their methods sections but with annotations retrieved from all sections. In general, any property that can be both measured and predicted can be stored, visualized, and analyzed similarly, as a chromatogram, image, or mass spectrum.

It is tempting to generalize the picture by traversing the ontology, looking at general classes rather than discrete chemical entities. However, two chemical entities in the same class (for example, methanol and octadecan-1-ol, both being primary alcohols) may have very different physicochemical properties. An alternative way to generalize the chemical entities associated with a particular method or process is to predict the most relevant properties of these entities using quantitative structure–property relationships (QSPR). When looking at multiple properties, the distributions can be visualized as heatmaps or images. Figure 3 shows the results from ALogPS 3.0 1-octanol–water-partition-coefficient (log *P*) QSPR predictions for all chemically distinct ChEBI entities, highlighting two classes with very different physicochemical properties.

Phase diagrams of applicability of analytical methods as functions of analyte mass and polarity are common in the literature,[25] including in textbooks, where they are used to compare methods of chromatographic separations or ionization in mass spectrometry. These diagrams are drawn from subjective experience rather than an objective metric. With the annotations in Europe PMC, it is now feasible to look at a large

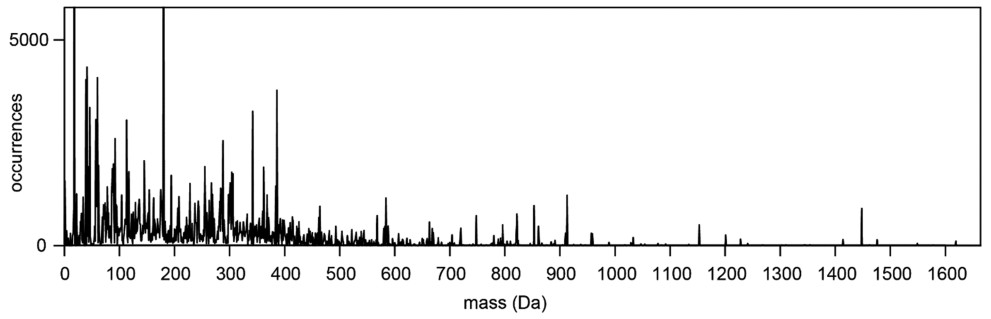


Figure 2. Distribution of integer masses of ChEBI compounds in publications with "liquid chromatography–mass spectrometry" in their methods sections visualized as an integer mass spectrum. The truncated base peak at mass 18 (18 491 occurrences) is dominated by ammonium ions and water. The second most abundant mass at 180 (14 497 occurrences) represents the hexoses. This simple analysis makes no distinctions among solvent, reagents, analytes, adducts, and neutral losses. However, analytes dominate at larger masses, and there are even discernible peaks for the anticancer drugs paclitaxel and docetaxel at 853 and 861 and the antibiotics sirolimus, colistin, cyclosporin A, and vancomycin at masses 913, 1150, 1201, and 1448, respectively. This is an actual mass spectrum, with mass rather than mass-to-charge ratio on the abscissa.

DOI: 10.1021/acs.analchem.8b05818
Anal. Chem. 2019, 91, 4312–4316

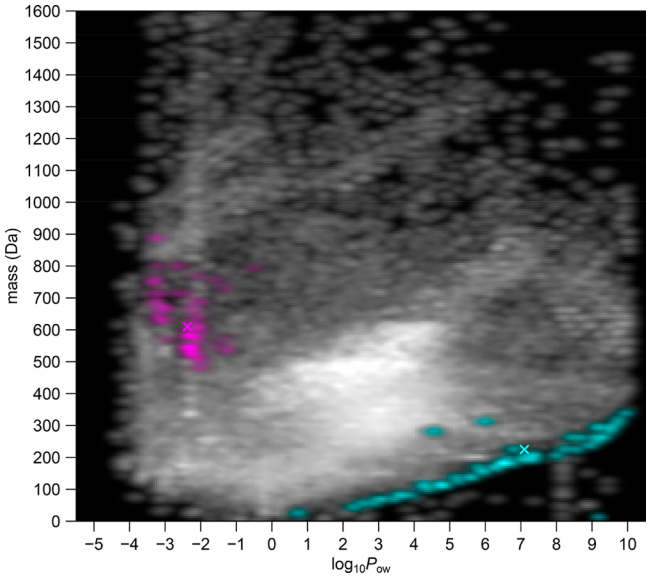


Figure 3. Prediction of 1-octanol–water-partition coefficients and masses for 90 056 ChEBI compounds. The instances of some ChEBI classes are clustered in the mass–log *P* space (e.g., the amino trisaccharides, CHEBI: 59266, magenta), whereas others, such as the alkanes (CHEBI: 18310, cyan), chart a predictable course through this mass–log *P* space. Although the ontological hierarchy provides a solid framework for systematically aggregating data on related compounds in a given context, it is important to remember the chemical diversity within an ontological class limits the representativeness of averages collated by traversing the ontology for all instances of this class in the literature on a particular topic. The crosses indicate the averages (or centers of mass) of the amino trisaccharides and alkanes in this map.

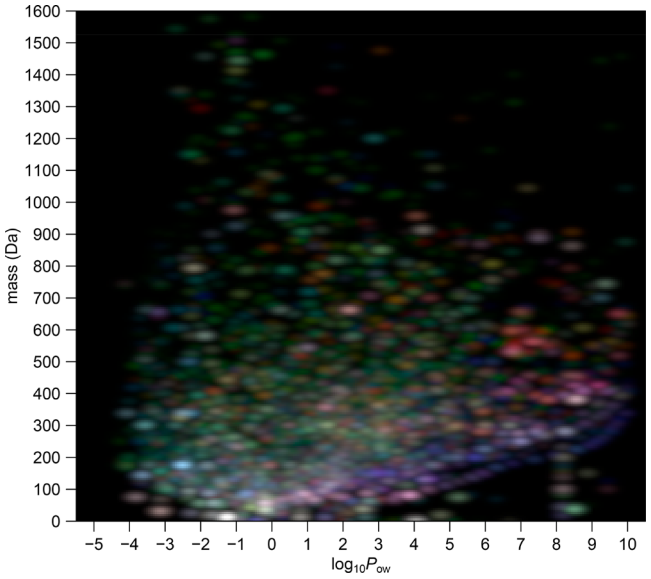


Figure 4. Normalized RGB plot[31] of the application of atmospheric-pressure chemical ionization–ionization (red), electrospray ionization–ionization (green), and electron impact (blue) from 1 623 096 named-entity recognitions in the scientific literature. Electron impact (now electron ionization) dominates for small compounds at the upper limit of log *P*, whereas APCI is the most popular ionization method for analytes near 500–600 Da and log *P* values of 6.5–8.5. A gray or white color indicates no preference among the three ionization methods, but this is only observed for ubiquitous solvents such as water.

body of the scientific literature and find *actual* correlations between analytical methods and analytes. Three common ionization methods for mass spectrometry were compared in a tripartite search for ESI, APCI, and electron impact (now called electron ionization), which retrieved 1 238 516, 116 868, and 267 712 ChEBI annotations from 36 844, 2567, and 5930 papers, respectively. The relative distributions are displayed in Figure 4. Electron impact (blue) follows the same distribution in polarity and mass as is observed for gas chromatography. This is unsurprising, as GC is most commonly interfaced to MS using this ionization method. Atmospheric-pressure chemical ionization has a distinct "sweet spot" at masses between 500 and 600 Da and 1-octanol–water-partition coefficients between $10^{6.5}$ and $10^{8.5}$. Electrospray dominates elsewhere in the mass–polarity space. The true phase diagram of ionization-method applicability is different from the rectangular diagrams found in the textbooks. However, it is important to consider that application depends on more than applicability, such as popularity and availability of particular instrumentation in the laboratories where published work is carried out. For example, LC-ESI-MS instrumentation is ubiquitous and versatile and provides appropriate choices of columns and mobile phases, which may explain why this combination has been so broadly applied to different types of analytes. Peptides and proteins are outside ChEBI's domain. As ChEBI contains fewer than 2000 peptides, and very few proteins, peptidomics, and proteomics data would not be efficiently captured even if the peptides were explicitly listed in the papers.

In addition to physicochemical properties and properties important in drug discovery, specific biological functions or interactions can be predicted by similar machine-learning techniques. From 78 630 predictions, 11 203 ChEBI compounds were predicted to be active estrogen-receptor (ER) agonists (numeric prediction <0.5; 0.323 ± 0.132, where 0 represents active, and 1 represents inactive) and 67 427 were predicted to be inactive (>0.5, 0.749 ± 0.111) with an 87.4% reported accuracy. A p53-signaling model predicted 28 718 out of 78 309 returned ChEBI compounds to be active (0.337 ± 0.114) and the remaining 49 591 to be inactive (0.722 ± 0.134) with 79.5% accuracy. The training of these predictors was performed using drugs and druglike compounds, whereas ChEBI contains many compounds that are neither. The results should therefore not be seen as reflecting poorly on any of the models, but as a useful way to indicate correlations between a particular topic or search term (e.g., an enzyme or receptor) and the small molecules coappearing in the literature with that enzyme or receptor.

The predicted most likely estrogen receptor (ER) agonists, not counting compounds used to train the model, were norgestrel (CHEBI: 7630), 17-ethynyl-13-methyl-7,8,9,11,12,14,15,16-octahydro-6*H*-cyclopenta[*a*]-phenanthrene-3,17-diol (CHEBI: 125402), (13*S*,17*R*)-17-ethynyl-13-methyl-7,8,9,11,12,14,15,16-octahydro-6*H*-cyclopenta[*a*]phenanthrene-3,17-diol (CHEBI: 91483), 1-(2,4-dihydroxyphenyl)-3-(3,4-dihydroxyphenyl)-2-propen-1-one (CHEBI: 92312), (8*R*,9*S*,13*S*,14*R*,17*R*)-17-ethynyl-13-methyl-7,8,9,11,12,14,15,16-octahydro-6*H*-cyclopenta[*a*]-phenanthrene-3,17-diol (CHEBI: 94792), and resveratrol (CHEBI: 27881). Several of these belong to the estrogen (CHEBI: 50114) class or are known ER agonists (resvera-

DOI: 10.1021/acs.analchem.8b05818
Anal. Chem. 2019, 91, 4312–4316

trol[26]), although (levo)norgestrel has relatively low affinity for the receptor.[27] The most likely p53-signaling agonists (excluding those in the training set) included more surprises, with C_{60} fullerene (CHEBI: 33128), quinacrine mustard (CHEBI: 37595), acridine half-mustard (CHEBI: 132980), pallidol (CHEBI: 27881), 2,2′,3′,4,4′,5,5′-heptachloro-3-biphenylol (CHEBI: 79726), 2,2′,3,3′,4′,5,5′-heptachloro-4-biphenylol (CHEBI: 34194), and 6-[[(6-methyl-9-indolo[3,2-*b*]quinoxalinyl)amino]methylidene]-1-cyclohexa-2,4-dienone (CHEBI: 94089) predicted to be the most likely active agonists, whereas known p53-signaling agonists in ChEBI (but not in the training set) such as piplartine−piperlongumine (CHEBI: 8241), 2,2-bis(hydroxymethyl)-1-azabicyclo[2.2.2]-octan-3-one (CHEBI: 94995), and the tripeptide acetylleucyl−leucyl−norleucinal (CHEBI: 2423) either failed prediction or were predicted to be inactive. These simple observations already suggest the ER predictions may be more accurate or at least more specific in the ChEBI domain than the p53-signaling predictions, despite the latter having a higher reported prediction accuracy. More relevantly, the 10 predicted most likely agonists returned by the Europe PMC literature searches for the estrogen receptor and cellular tumor antigen p53 are shown in Table 1. The list for the former included the estrogens (CHEBI: 50114) 17β-estradiol, estradiol, and estrone, as well as the xenoestrogen (CHEBI: 76988) dienestrol and the androgen (CHEBI: 50113) testosterone; the list for the latter included the antineoplastic agents (CHEBI: 35610) torin 1, nocodazole, and irinotecan but also the fluorochromes (CHEBI: 51217) methylene blue, pibenzimol, ethidium, thionine, and 2′-(4-ethoxyphenyl)-5-(4-methylpiperazin-1-yl)-2,5′-bibenzimidazole (a pibenzimol derivative). Although these fluorochromes all bind to DNA, and several have been investigated for use in photodynamic anticancer therapy,[28,29] they are most often used for staining, also in the context of p53-signaling studies. Nevertheless, these results clearly demonstrate the feasibility of semantically enhanced literature searches combining named-entity recognition of proteins and their ligands from a small-molecule ontology with large-scale QSPR predictions.

Table 1. Most Likely ER and p53-Signaling Agonists As Predicted by OCHEM Matching the Respective Protein in a Europe PMC Search

top ER agonists in {UNIPROT_PUBS:P03372}		
ChEBI ID	ChEBI name	number prediction
16469	17β-estradiol	2.67×10^{-03}
23965	estradiol	3.23×10^{-03}
34025	1,1,1-trichloro-2,2-bis(4-hydroxyphenyl)ethane	1.13×10^{-02}
57545	2-(3,4-dihydroxyphenyl)-5-hydroxy-4-oxo-4*H*-chromen-7-olate luteolin-7-olate(1−)	1.20×10^{-02}
17347	testosterone	1.97×10^{-02}
31669	hexestrol	2.17×10^{-02}
3908	coumestrol	3.18×10^{-02}
1156	2-hydroxyestrone	3.80×10^{-02}
4518	dienestrol	3.87×10^{-02}
17263	estrone	4.17×10^{-02}
top p53-signaling agonists in {UNIPROT_PUBS:P04637}		
ChEBI ID	ChEBI name	number prediction
51739	acridine orange	1.37×10^{-02}
6872	methylene blue	1.86×10^{-02}
52082	pibenzimol	2.07×10^{-02}
84327	torin 1	2.43×10^{-02}
4883	ethidium bromide	3.09×10^{-02}
42478	ethidium	3.09×10^{-02}
52295	thionine	3.30×10^{-02}
34892	nocodazole	3.83×10^{-02}
51232	2′-(4-ethoxyphenyl)-5-(4-methylpiperazin-1-yl)-2,5′-bibenzimidazole	3.87×10^{-02}
80630	irinotecan	4.67×10^{-02}

DISCUSSION

The inspiration for the method described here was to demonstrate the power of combining existing resources and web services for novel purposes. The number of ways in which literature databases, ontologies, text-mined annotations, and QSPR modeling can complement each other is inexhaustible. Multiple ontologies can be combined, such as protein entities in UniProt (enzymes and receptors), small molecules in ChEBI (as demonstrated here), and also chemical-method ontologies like CHMO.[30] Entities from any of the supported ontologies can be used in the initial search query in Europe PMC. Much of the chemistry data and software is proprietary, and the free software and services that exist typically only support single queries or a small number of queries. Here, gratis data, software, and services that are scalable to millions of queries or predictions were used exclusively, meaning anyone with an Internet connection can repeat or modify the analyses using the workflow provided in the GitHub repository.

In these demonstrations, no distinction was made between proper analytes and reagents used in the analysis. Automatically tagging these requires AI-hard natural-language understanding. The shortcut can be defended by the safe assumption that for almost any analytical method or literature search, the number of reported analytes will be much larger than the number of different additives and derivatizing agents used in the method. Comparing techniques for analysis or synthesis using text mining and molecular-property calculations not only summarizes what has been done but inspire selection of existing methods or development of new methods that are fit-for-purpose. The examples here were primarily chosen as positive controls, to check that the resulting visualizations are reasonable and informative.

In principle, all analyses here could just as easily be performed on a larger set of compounds, provided these would be annotated in the literature. Annotating PubMed with PubChem, containing 97 million compounds as of March 2019, is an obvious next step. Some differences are expected, as the coverage of PubChem and ChEBI differ (by design). Compared with PubChem, ChEBI is heavily influenced by endogenous metabolites and other large, water-soluble molecules, resulting in the average ChEBI compound being significantly larger (~440 vs ~370 Da) and more polar (log *P* ~2.27 vs ~2.89) than the average of the 97 million PubChem compounds calculated using the ALogPS 3.0 model. None of these differences should pose any technical difficulties, except for the most computationally expensive models.

CONCLUSIONS

This note describes a novel method integrating existing resources for literature searches, text mining, and computational chemistry. The method has potential applicability in

DOI: 10.1021/acs.analchem.8b05818
Anal. Chem. 2019, 91, 4312−4316

experimental design, drug development, omics-data integration, and historical studies of scientific fields. It may also be used for business intelligence, characterizing strengths and weaknesses, and identifying potential opportunities by comparing products from competing vendors. Linking meaningful categories, functions, or quantitative properties to text-mined items semantically enhances the literature searches, and their aggregation and visualization provide in a single picture an overview of an analytical technique or methodology.

AUTHOR INFORMATION

Corresponding Author

*E-mail: magnus.palmblad@gmail.com.

ORCID

Magnus Palmblad: 0000-0002-5865-8994

Notes

The author declares no competing financial interest.

ACKNOWLEDGMENTS

The author wishes to acknowledge Dr. Igor V. Tetko at the Institute of Structural Biology, Helmholtz Zentrum München, for his generous assistance with the OCHEM server; Dr. Aswin Verhoeven for sharing his KNIME expertise; and Dr. Rico J. E. Derks for fruitful discussions on QPSR.

REFERENCES

(1) Europe PMC Consortium. *Nucleic Acids Res.* **2015**, *43*, D1042−D1048.

(2) Venkatesan, A.; Kim, J. H.; Talo, F.; Ide-Smith, M.; Gobeill, J.; Carter, J.; Batista-Navarro, R.; Ananiadou, S.; Ruch, P.; McEntyre, J. *Wellcome Open Res.* **2016**, *1*, 25.

(3) Guler, A. T.; Waaijer, C. J.; Palmblad, M. *Scientometrics* **2016**, *107*, 385−398.

(4) Guler, A. T.; Waaijer, C. J. F.; Mohammed, Y.; Palmblad, M. *J. Informetr* **2016**, *10* (3), 830−841.

(5) Palmblad, M.; Torvik, V. I. *Trop Med. Health* **2017**, *45*, 33.

(6) Wolstencroft, K.; Haines, R.; Fellows, D.; Williams, A.; Withers, D.; Owen, S.; Soiland-Reyes, S.; Dunlop, I.; Nenadic, A.; Fisher, P.; Bhagat, J.; Belhajjame, K.; Bacall, F.; Hardisty, A.; Nieva de la Hidalga, A.; Balcazar Vargas, M. P.; Sufi, S.; Goble, C. *Nucleic Acids Res.* **2013**, *41*, W557−W561.

(7) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B., KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications. Studies in Classification, Data Analysis, and Knowledge Organization*; Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R., Eds.; Springer: Berlin, 2008.

(8) Hastings, J.; de Matos, P.; Dekker, A.; Ennis, M.; Harsha, B.; Kale, N.; Muthukrishnan, V.; Owen, G.; Turner, S.; Williams, M.; Steinbeck, C. *Nucleic Acids Res.* **2012**, *41*, D456−D463.

(9) Anderson, E.; Veith, G. D.; Weininger, D. *SMILES: A line notation and computerized interpreter for chemical structures*; U.S. EPA, Environmental Research Laboratory: Duluth, MN, 1987.

(10) Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I. *J. Cheminf.* **2013**, *5* (1), 7.

(11) McDonagh, J. L.; van Mourik, T.; Mitchell, J. B. *Mol. Inf.* **2015**, *34* (11−12), 715−724.

(12) Hewitt, M.; Cronin, M. T. D.; Enoch, S. J.; Madden, J. C.; Roberts, D. W.; Dearden, J. C. *J. Chem. Inf. Model.* **2009**, *49* (11), 2572−2587.

(13) Raevsky, O. A.; Polianczyk, D. E.; Grigorev, V. Y.; Raevskaja, O. E.; Dearden, J. C. *Mol. Inf.* **2015**, *34* (6−7), 417−430.

(14) Tetko, I. V.; Sushko, Y.; Novotarskyi, S.; Patiny, L.; Kondratov, I.; Petrenko, A. E.; Charochkina, L.; Asiri, A. M. *J. Chem. Inf. Model.* **2014**, *54* (12), 3320−3329.

(15) Wang, L. H.; Hsieh, C. M.; Lin, S. T. *Ind. Eng. Chem. Res.* **2015**, *54* (41), 10115−10125.

(16) Vetere, A. *Fluid Phase Equilib.* **1991**, *62* (1−2), 1−10.

(17) Benet, L. Z.; Broccatelli, F.; Oprea, T. I. *AAPS J.* **2011**, *13* (4), 519−547.

(18) Daina, A.; Michielin, O.; Zoete, V. *Sci. Rep.* **2017**, *7*, 42717.

(19) Zhang, H.; Ren, J. X.; Kang, Y. L.; Bo, P.; Liang, J. Y.; Ding, L.; Kong, W. B.; Zhang, J. *Reprod. Toxicol.* **2017**, *71*, 8−15.

(20) Marzo, M.; Roncaglioni, A.; Kulkarni, S.; Barton-Maclaren, T. S.; Benfenati, E. *Methods Mol. Biol.* **2016**, *1425*, 139−161.

(21) Abdelaziz, A.; Spahn-Langguth, H.; Schramm, K.-W.; Tetko, I. V. *Front. Environ. Sci.* **2016**, *4*, 2.

(22) Kafkas, S.; Pi, X.; Marinos, N.; Talo', F.; Morrison, A.; McEntyre, J. R. *J. Biomed Semantics* **2015**, *6*, 7.

(23) Sushko, I.; Novotarskyi, S.; Korner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. V.; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-de-Sousa, J.; Zhang, Q. Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. *J. Comput.-Aided Mol. Des.* **2011**, *25* (6), 533−554.

(24) Kavlock, R. J.; Austin, C. P.; Tice, R. R. *Risk Anal* **2009**, *29* (4), 485−487.

(25) Gu, C.; Lin, B.; Pease, J.; Chetwyn, N.; Yehl, P. Mass Spectrometry in Small Molecule Drug Development. *Am. Pharm. Rev.*, Sept 30, 2015.

(26) Levenson, A. S.; Gehm, B. D.; Pearce, S. T.; Horiguchi, J.; Simons, L. A.; Ward, J. E., 3rd; Jameson, J. L.; Jordan, V. C. *Int. J. Cancer* **2003**, *104* (5), 587−596.

(27) Sitruk-Ware, R. *Hum. Reprod. Update* **2006**, *12* (2), 169−178.

(28) dos Santos, A. F.; Terra, L. F.; Wailemann, R. A. M.; Oliveira, T. C.; Gomes, V. D.; Mineiro, M. F.; Meotti, F. C.; Bruni-Cardoso, A.; Baptista, M. S.; Labriola, L. *BMC Cancer* **2017**, *17* (1), 194.

(29) Fowler, G. J.; Rees, R. C.; Devonshire, R. *Photochem. Photobiol.* **1990**, *52* (3), 489−494.

(30) Batchelor, C. The Chemical Methods Ontology, 2016. *GitHub*. https://github.com/rsc-ontologies/rsc-cmo (accessed Feb 05, 2019).

(31) Suari, Y.; Brenner, S. *PLoS One* **2014**, *9* (7), No. e102903.

DOI: 10.1021/acs.analchem.8b05818
Anal. Chem. 2019, 91, 4312−4316