JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

Editorial

What Makes a Paper Be Highly Cited? 60 Years of the *Journal of Chemical Information and Modeling*

Cite This: *J. Chem. Inf. Model.* 2020, 60, 5866−5867

Read Online

ACCESS | Metrics & More | Article Recommendations

In 2020, the *Journal of Chemical Information and Modeling* (JCIM), the premier journal for research in cheminformatics and modeling, celebrates the 60th anniversary of its founding. Over the years, this peer-reviewed journal published by the American Chemical Society has followed the birth and development of an emerging new field that encompasses chemistry, biology, and computer science. Together with the evolution of the new discipline, its name has changed three times: it was established in 1961 as the *Journal of Chemical Documentation* (JDC), which was later renamed in 1975 as the *Journal of Chemical Information and Computer Sciences* (JCICS). In 2005, the current name of *Journal of Chemical Information and Modeling* was adopted. Since its first edition, it stands out as a core journal covering the broad field that is computational chemistry. For example, many of the most influential discoveries that marked milestones in chemoinformatics were shared with the scientific community by JCICS or JCIM. In this editorial, we celebrate the importance of JCIM in computational chemistry showcasing 11 papers, selected among the most cited ones published in the last six decades. These papers, chosen by members of the Early Career Board of JCIM, are part of a Virtual Issue to celebrate JCIM's anniversary. They cover a significant time span, from 1988 to 2015, and were authored by researchers from different countries, such as the United States, China, Austria, and the United Kingdom. In what follows, we briefly describe these articles and analyze the reasons behind their success.

Free databases. In the era of Big Data, where experiments such as gene sequencing and high-throughput assays generate a very large amount of information, it is useful to have an organized data collection. Databases are well suited for this purpose. They are important for storing information in a comprehensive manner, and they also facilitate the building of data sets for data mining and training of computational methods. The accuracy of the data used for training has a huge influence on the models obtained, since the models are only as good as the data.[1] In this Virtual Issue, we find four papers related to freely available databases.[2−5] Three of them are related to ZINC,[2−4] which is a database of commercially available compounds for virtual screening created by John Irwin and Brian Shoichet in 2005. It contains over 750 million small molecules with annotated properties, such as molecular weight and LogP, and structures ready to use in docking. Reference 2 is the most cited paper in JCIM so far, with more than 3200 citations, which shows how widespread the use of ZINC is. The other paper describes admetSAR,[5] which is a database created by Yun Tang and colleagues. It hosts a collection of curated data from the literature regarding absorption, distribution, metabolism, excretion, and toxicity (ADMET) of small molecules. These databases are invaluable tools for rational drug design.

Free tools. Free tools are fundamental to speed up research, both in the context of companies and (even more) for academia. The main advantages are the cost reduction of the project, as well as the acceleration of its development by having fewer restrictions than those using closed models. When the code is open source, another advantage is the possibility of code maintenance and improvement by other members of the community, as pointed out in a recent Viewpoint.[6] In 2012, two well-known free tools, which are constantly highly cited, were published in JCIM.[5,7,8] The first regards the prediction of some molecular features. It describes admetSAR[5] which, beyond being a database of ADMET data, also includes models to predict pharmacokinetic properties of new molecules. Metabolism prediction plays a crucial role in reducing attrition during drug discovery, and this useful tool can be exploited to quickly integrate drug-design studies involving potential active molecules with the prediction of their ADMET properties. The second example of a successful free tool deals with the world of molecular mechanics simulations applied to computer aided drug design. Alexander MacKerell and colleagues developed a web server for automatic atom typing that generates parameters for small molecules.[7,8] It generates CHARMM general force field (CGenFF) models, which is the corresponding force field for drug-like molecules to the well-established CHARMM force field for biomolecular systems. It facilitates accurate molecular dynamics-based studies, since it is fast and easy to use.

Tools for drug design. Computational methods are useful tools in rational drug design. For instance, they can identify important ligand-protein interactions that can lead to improved affinity, or they can be used to predict properties such as toxicity or absorption. These methods, however, are still far from perfect and usually require iterations between predictions and experiments to be optimized. In addition to the free databases and tools already mentioned, two highly

Published: December 28, 2020

https://dx.doi.org/10.1021/acs.jcim.0c01248
J. Chem. Inf. Model. 2020, 60, 5866−5867

cited papers describe tools for rational drug design.[9,10] One of the tools is LigandScout,[9] which was developed by Gerhard Wolber and Thierry Langer to create pharmacophore models to be used in virtual screening. The rules to detect interactions to generate the models were obtained by data mining protein−ligand complexes in the Protein Data Bank (PDB). The other tool is OMEGA,[10] which was created by Paul Hawkins and colleagues to generate conformers for small molecules. It has been validated using structures of small molecules from the PDB and the Cambridge Structural Database. Exhaustive conformer generation for small molecules and accurate pharmacophore modeling are still challenging tasks in drug discovery.

Comparison of methods. A comparison of different methods is valuable to understand the strengths and weaknesses of each method and make recommendations regarding which tool works best for specific cases. Two papers featured in the Virtual Issue are related to method comparison.[11,12] In one of them, Tingjun Hou and colleagues compare the performance of Molecular Mechanics/Poisson−Boltzmann Surface Area (MM/PBSA) and Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) in the prediction of relative and absolute binding free energies for 59 ligand-protein complexes.[11] The authors also evaluate the impact of other factors in the predictions, such as simulation length and solute dielectric constant. In the other paper, Peter Willett and colleagues review methods of similarity searching in chemical databases.[12] The paper showcases methods to quantify the similarity between molecules and to represent molecular structures, and it finishes by comparing different approaches for similarity searching. Effective ways to represent molecules, which keep structural information to facilitate similarity searches, are still an interesting research topic in computational chemistry.

Innovative ideas. Finally, new ideas to represent information can have a long-term impact on the field. In one of the papers highlighted in the Virtual Issue, David Weininger describes the creation of SMILES, a chemical language to represent small molecules using a linear string of characters and symbols.[13] Its use is widespread nowadays, and every paper in the area of computational chemistry dealing with small molecules usually describes them using SMILES to guarantee transparency and reproducibility. Reference 13 is the oldest paper in the Virtual Issue, published in 1988, and also the most cited one in JCICS, with more than 3000 citations.

What will the highly cited papers look like in the coming years? Some of the past trends will remain. Papers describing free databases with curated and quality data will keep being heavily cited, given the widespread use of machine learning and methods for nonlinear fitting. Papers describing free tools in drug design will still be popular, especially the ones dealing with unmet challenges, such as prediction of hydration sites in protein pockets, covalent docking, and design of new synthesizable molecules. Moreover, we also expect that papers with methods and algorithms for computational chemistry designed to have improved performance for new technologies, such as graphics processing units (GPU), tensor processing units (TPU), cloud computing, and quantum computing, will be highly cited due to the widespread use of these tools.

Ariane Nunes-Alves orcid.org/0000-0002-5488-4732
Angelica Mazzolari orcid.org/0000-0003-1352-1094
Kenneth M. Merz, Jr. orcid.org/0000-0001-9139-5893

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.0c01248

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

REFERENCES

(1) Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model.* **2016**, *56* (7), 1243−1252.
(2) Irwin, J. J.; Shoichet, B. K. ZINC − A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177−182.
(3) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52* (7), 1757−1768.
(4) Sterling, T.; Irwin, J. J. ZINC 15 − Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55* (11), 2324−2337.
(5) Cheng, F.; Li, W.; Zhou, Y.; Shen, J.; Wu, Z.; Liu, G.; Lee, P. W.; Tang, Y. AdmetSAR: A Comprehensive Source and Free Tool for Assessment of Chemical ADMET Properties. *J. Chem. Inf. Model.* **2012**, *52* (11), 3099−3105.
(6) Walters, W. P. Code Sharing in the Open Science Era. *J. Chem. Inf. Model.* **2020**, *60* (10), 4417−4420.
(7) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *J. Chem. Inf. Model.* **2012**, *52* (12), 3144−3154.
(8) Vanommeslaeghe, K.; Raman, E. P.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *J. Chem. Inf. Model.* **2012**, *52* (12), 3155−3168.
(9) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45* (1), 160−169.
(10) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50* (4), 572−584.
(11) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2011**, *51* (1), 69−82.
(12) Willett, P.; Barnard, J. M.; Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38* (6), 983−996.
(13) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28* (1), 31−36.

https://dx.doi.org/10.1021/acs.jcim.0c01248
J. Chem. Inf. Model. 2020, 60, 5866−5867