

SOFTWARE

Open Access



Twitter integration of chemistry software tools

Naruki Yoshikawa^{1†}, Ryuichi Kubo^{3†} and Kazuki Z. Yamamoto^{2,3,4*}

Abstract

Social media activity on a research article is considered to be an altmetric, a new measure to estimate research impact. Demonstrating software on Twitter is a powerful way to attract attention from a larger audience. Twitter integration of software can also lower the barriers to trying the tools and make it easier to save and share the output. We present three case studies of Twitter bots for cheminformatics: retrosynthetic analysis, 3D molecule viewer, and 2D chemical structure editor. These bots make software research more accessible to a broader range of people and facilitate the sharing of chemical knowledge, concepts, and ideas.

Keywords: Twitter, Social media, Retrosynthesis, Visualization, Molecule editor, Chemical space, Knowledge sharing, Open science

Introduction

Social media such as Twitter has become an important communication tool for scientists. *Twitter* is a social networking service on which users post short messages called *tweet*, interact with each other by sending a *reply* and share other users' tweets by *retweet*. Scientists are using *hashtags* such as #chemtwitter, #RealTimeChem, #ScienceTwitter, #compchem and #OpenScience to create communities on Twitter for networking and information exchange [1]. Twitter activities about a research paper are considered to be one of the altmetrics [2], which is an alternative to citation counting for measuring scientific impact. Some studies reported that the number of tweets on a research article is positively correlated to the article citations [3–5]. Attracting attention on Twitter is getting more important to promote scientific works.

A Twitter bot is an automated program on Twitter controlled through Twitter API. Bots are common on Twitter, and they usually deliver tweets like news and blog

updates [6]. Twitter bots are also used for scientific purposes, such as posting updates of journal publications or preprints. While most bots automatically post tweets, they can also respond to replies from other users. We can construct an interactive application using Twitter API as long as following the rules and policies of Twitter such as Automation rules [7].

In this paper, we report case studies of cheminformatics tools demonstrated on Twitter by introducing the functionality of our Twitter bots and Twitter cards. The Twitter account @souyakuchan [8] was originally created for promoting the authors' project of crowd-based drug discovery (named Social Drug Discovery). It has been hosting a simple cheminformatics bot since 2017 in the form of a bot that draws a 2D structural formula image for replies containing a SMILES or an IUPAC name [9]. The bot was later expanded by integrating functions in `chemical.space` [10], which are described in the case study 2 and 3. The second account @retrosynthchan [11] offers retrosynthesis functionality. It was implemented in December 2020 for the blog event called Drug Discovery (dry) Advent Calendar 2020 [12] hosted by @souyakuchan. We describe the retrosynthesis bot in the case study 1.

*Correspondence: kazuki@ric.u-tokyo.ac.jp

[†]Naruki Yoshikawa and Ryuichi Kubo contributed equally to this work

² Isotope Science Center, The University of Tokyo, Tokyo, Japan
Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Case study 1: retrosynthesis bot

In this section, we introduce a Twitter bot @retrosynthchan [11], which conducts retrosynthetic analysis by AiZynthFinder [13]. The bot responds to a request given as a SMILES [14] string and returns the results of retrosynthetic analysis for the given molecule. Users are expected to send a reply containing only a SMILES string right after the bot's screen name (e.g., @retrosynthchan c1ccccc1Br). The communication between users and the bot is in public, so every Twitter user can see the query and result.

Implementation

The workflow of the bot program is as follows: (1) retrieve a reply to the bot via Twitter API, (2) check whether the given string is valid as SMILES using RDKit [15] and send a request confirmation, (3) run retrosynthesis analysis by AiZynthFinder [13], (4) send the retrosynthesis result images via Twitter API. The details of each step will be discussed.

Retrieve reply

A reply in Twitter is a message toward a specific account, and it usually starts with the screen name of the receiver (e.g., @jack How are you today?). At the time of this writing, the Twitter API provides a method to retrieve tweets that contain a specific keyword in real-time (`POST statuses/filter` [16]). The bot receives a reply to the bot by retrieving tweets that contain the bot's screen name with this API. A Python library Tweepy [17] is used to handle Twitter API.

Confirm request

After receiving a reply, the string is separated into parts by a space. The second part is assumed to be a SMILES string since the first part is the bot's screen name. The SMILES string is considered to be valid if it can be successfully parsed by RDKit [15] with the default setting. The bot returns the input molecule's chemical structure if the input is valid, otherwise, send back an error message to the sender and finish. The confirmation is sent as a reply to the users by a Twitter API `POST statuses/update` [18] using Tweepy.

Run retrosynthetic analysis

Retrosynthetic analysis of the molecule represented by the given SMILES is conducted by AiZynthFinder [13], open-source software for retrosynthetic planning. The input molecule is broken down into purchasable precursors with the algorithm based on a Monte Carlo tree

search guided by a neural network. The synthesis route images are generated for individual routes.

Send synthetic pathway images

The generated synthetic route images are sent to the user as a reply with images. Twitter allows users to attach only up to four images in one post, so three images are connected into one image using Pillow [19] to send a maximum of twelve synthesis routes. The generated images are sent back to the users with the Twitter API using Tweepy.

Results

A screenshot of the bot in action is shown in Fig. 1. A request confirmation with a chemical structure image was sent right after the SMILES was sent to the bot. The result of the retrosynthetic analysis is sent several minutes later since it takes a while to run AiZynthFinder. Typically, it takes less than 3 min to return the result since we use the default time limit (120 s) of AiZynthFinder, but it may be delayed if other people are running the job.

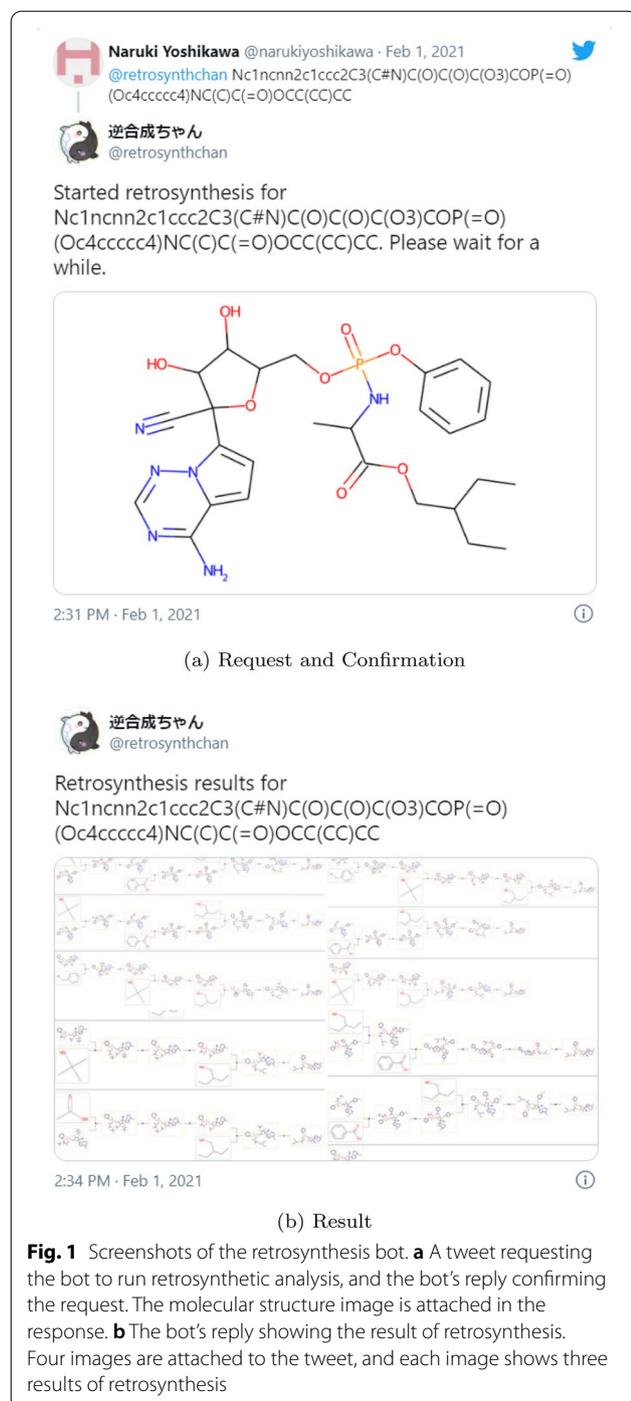
The bot attracted attention from Twitter users interested in chemistry, and it gained more than one hundred followers in the first 2 weeks. The followers include chemists in academia, researchers in the drug industry, undergraduate or graduate chemistry students, etc. The Twitter bot helped the retrosynthesis software outreach to a broader range of people.

Case study 2: 3D molecule viewer embedded in Twitter timeline

Next, we introduce a 3D molecule viewer which can be played on a Twitter timeline inlined via our web server `chemical.space` [10]. This feature is based on the Twitter Player Card system [20]. The Player Card system is used as a video/audio player in the Twitter timeline, wherein the 3D molecule viewer works as interactive video content. This feature is currently available on web browsers, while the Twitter mobile application can not load Player Card directly in the timeline due to the limitation of Twitter's specifications. Users can view any Protein Data Bank (PDB) structures interactively without leaving their timeline on the browser. To embed the viewer in a tweet, users are expected to post a tweet containing a PDB ID string right after the bot's screen name with a #pdb hashtag (e.g., @souyakuchan lema #pdb).

Implementation

As a 3D molecule viewer, we adopted LiteMol [21], which is a JavaScript application written in TypeScript. The way of receiving requests is the same as in Case 1.



After receiving the PDB ID via users' tweets, the bot makes a reply in which the URL of our hosting server of LiteMol is indicated. LiteMol refers to the URL parameter and fetches the PDB data. Also, to provide Twitter Player Card, our viewer web page serves meta information in HTML header named `twitter:card`. Twitter Player Card can render HTML referring an URL in the

tweet. The meta information is mainly utilized for recognizing the player (embedded web page) URL and the player title.

Results

A screenshot of the viewer played on a Twitter timeline is shown in Fig. 2. After receiving a PDB code by a *mention* tweet, the bot immediately replies with an URL in which the 3D viewer Player Card can be expanded. Users can view the 3D structure interactively as it is in the Twitter timeline.

Case study 3: shareable 2D chemical structural formula editor

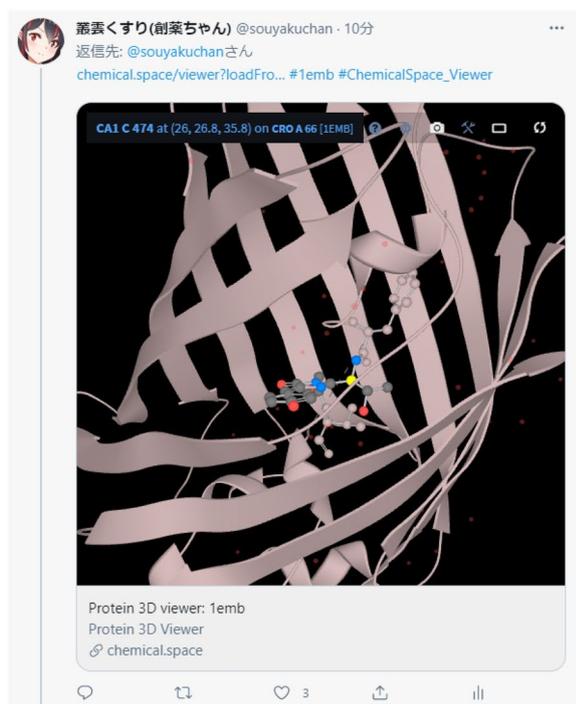
Finally, we introduce a Twitter Player Card which offers an easy way to share a molecule inside a tweet. The 2D chemical structural formula editor integrated with tweet posting function (available at <https://chemical.space/editor>) is one of the easiest ways to share one's own idea of chemical structures on Twitter. Users can share their manually edited structural formula as a SMILES or Chemical Markup Language (CML) [22] format. By using CML, users can include the information of coordinate and rotation of each object. This is a better way to share complex structures, whereas SMILES can only represent the molecular graph information, which has the advantage of the smallest data size. This editor has one more useful feature, in which users can send their manually drawn chemical structure to @retrosynthchan bot (shown in Case Study 1) by just clicking the RETROSYNTHESIZE button.

Implementation

Among the existing well-designed traditional browser-based 2D molecule editors such as JSME [23] and MarvinSketch [24], we chose Kekule.js [25] for its ease of integration as a web application and its sufficient functionality. Its JavaScript-based editor can input and output SMILES strings and CML strings of manually drawn structural formulas. Also, we used rdkit.js [26] to convert SMILES query to Molfile format [27] containing 3D coordinates. While SMILES only stores simple characters to express atoms, bonds, rings, and their orders, CML has coordinate and rotation information of each atom and bond. Furthermore, CML contains many XML markup tags in addition to that information. Meanwhile, CML strings become too long to tweet because one tweet accepts only 280 characters. However, URL strings are not included in this limitation. To embed CML strings into an URL, <https://chemical.space/editor> compresses

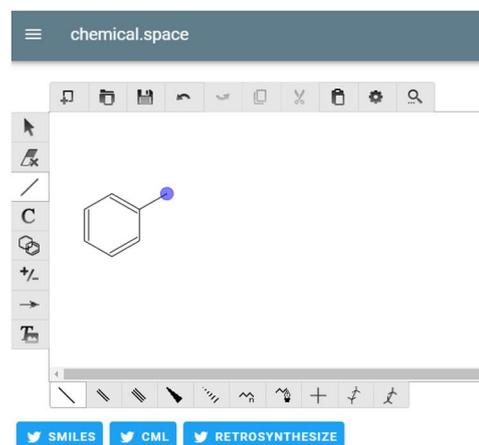


(a) Request and reply

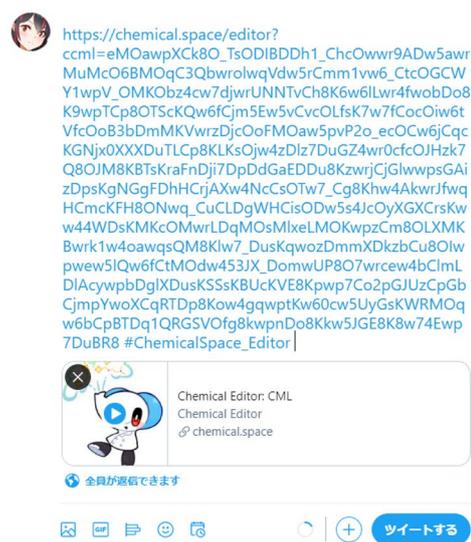


(b) View

Fig. 2 Screenshots of the Twitter integrated 3D molecule viewer. **a** A tweet requesting the bot to return an URL of a 3D molecular viewer and the subsequent reply with the viewer embedded in the Twitter timeline. Users can expand the viewer by clicking the play button. **b** The viewer expanded in the Twitter timeline. Users can interactively manipulate the view



(a) Edit



(b) Share



(c) Display

Fig. 3 Screenshots of the Twitter-integrated 2D editor. **a** The 2D molecule editor integrated with tweet buttons. **b** The tweet input window to share the manually drawn structural formula. The CML strings are encoded in the URL. **c** The 2D molecule viewer expanded on the Twitter timeline. At the moment, intra-timeline player is only supported on the PC browser version of Twitter, and not on the mobile App version

CML strings by gzip using pako [28] library, then encodes it into Base64url using brianloveswords/base64url [29].

Results

The 2D molecule editor hosted on <https://chemical.space/editor> is shown in Fig. 3. Users can utilize it to draw structural formulas from scratch in their browser and post them to Twitter in SMILES or CML format using the button at the bottom of the editor. We have also prepared a button that allows users to submit their structural formula to the retrosynthetic analysis bot introduced in Case Study 1. Structural formulas posted in CML format can be re-edited by opening the URL in the tweet, which is useful for open discussion on compounds on Twitter.

Discussion

As presented in these case studies, it is technically possible to make any cheminformatics tool accessible and noticeable from social networking services in the form of interfacing with Twitter bots. However, to make effective impressions on the scientific community and stimulate discussion, it is necessary to use appropriate hashtags that are widely prevailing and be tweeted by accounts with a large number of followers and strong influence. It would be helpful for scientists to follow accounts of mutual interest regularly and grow their follow/follower count appropriately. Tweets with high-impact and sensational content are likely to generate more impressions, but the information described in the tweet must have a certain level of accuracy. As with the sanitization of structural formulas [30], it is also beneficial for peers of scientists to ensure each other's scientific soundness on the Twitter timeline. An attempt to socially share ideas for compounds has also emerged very recently in the fight against the pandemic coronaviral disease COVID-19 (e.g., the COVID Moonshot project [31–33]), which is an example of the trend toward open science in drug discovery. Looking ahead to applications in drug discovery, we also need widgets in which users can interactively view and discuss the interaction and activity of compounds with target proteins, which is a future challenge. As a semi-permanently maintained database infrastructure, Twitter has the potential to store an unlimited amount of chemical information.

Conclusions

We presented three examples of Twitter integrated cheminformatics tools in this study. Twitter integration can offer a convenient interface to run cheminformatics software without a laborious environment setup, and

users can easily share the result by tweet and retweet. Providing a demonstration on social media will be a new approach to present one's research and stimulate discussion in a more accessible manner. If you would like to add your own tool to our bot functions, please contact the authors for support.

Availability and requirements

Case study 1

- Project name: Twitter Retrosynthesis Bot
- Project home page: https://github.com/n-yoshikawa/retrosynthesis_bot
- Operating system(s): Platform independent
- Programming languages: Python
- Other requirements: AiZynthFinder, RDKit, Tweepy, Pillow
- License: MIT License

Case study 2 and 3

- Project name: chemical.space
- Project home page: <https://github.com/ddquest/chemical.space>
- Operating system(s): Platform independent
- Programming languages: Python
- Other requirements: LiteMol, RDKit, Tweepy
- License: MIT License

Acknowledgements

Thanks to all our followers who interact with us on Twitter. Masahito Ohue and Mario Krenn provided helpful suggestions for this manuscript.

Disclaimer

TWITTER, TWEET, RETWEET and the Twitter Bird logo are trademarks of Twitter Inc. or its affiliates. We followed Twitter Trademark Guidelines, Terms of Service and all other Twitter rules and policies in this research and manuscript to the best of our knowledge.

Authors' contributions

NY implemented the software for case study 1, and RK implemented the software for case study 2 and 3. NY, RK, and KZY wrote the paper. KZY conceived the outline of `chemical.space` project. KZY maintains the cloud web servers. All authors read and approved the final manuscript.

Funding

This work was supported by JSPS KAKENHI Grant Number 18K11523. The server maintenance cost has been and will be paid by SHaLX Inc.

Availability of data and materials

The source codes of Case 1 software are available at <https://doi.org/10.5281/zenodo.4999041>. The source codes of the Case 2 and 3 software are available at <https://doi.org/10.5281/zenodo.4979985>. The `chemical.space` applications are available at <https://chemical.space/>.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Computer Science, University of Toronto, Toronto, Canada. ²Isotope Science Center, The University of Tokyo, Tokyo, Japan. ³SHaLX Inc., Tokyo, Japan. ⁴Department of Computer Science, Tokyo Institute of Technology, Yokohama, Japan.

Received: 13 April 2021 Accepted: 15 June 2021

Published online: 02 July 2021

References

- Reeser D, Vincent-Ruz P, Ashwell S, Kemsley J, Yarnell A (2021) The secret silos of #ChemTwitter. <https://cen.acs.org/sections/the-secret-silos-of-chemtwitter.html>. Accessed 19 Mar 2021
- Priem J, Groth P, Taraborelli D (2012) The altmetrics collection. *PLoS ONE* 7(11):48753. <https://doi.org/10.1371/journal.pone.0048753>
- Eysenbach G (2011) Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res* 13(4):123. <https://doi.org/10.2196/jmir.2012>
- Peoples BK, Midway SR, Sackett D, Lynch A, Cooney PB (2016) Twitter predicts citation rates of ecological research. *PLoS ONE* 11(11):0166570. <https://doi.org/10.1371/journal.pone.0166570>
- Klar S, Krupnikov Y, Ryan JB, Searles K, Shmargad Y (2020) Using social media to promote academic research: identifying the benefits of twitter for sharing academic work. *PLoS ONE* 15(4):0229446. <https://doi.org/10.1371/journal.pone.0229446>
- Chu Z, Gianvecchio S, Wang H, Jajodia S (2010) Who is tweeting on Twitter: human, bot, or cyborg? In: Proceedings of the 26th annual computer security applications conference. pp 21–30. <https://doi.org/10.1145/1920261.1920265>
- Twitter's automation development rules | Twitter Help. <https://help.twitter.com/en/rules-and-policies/twitter-automation>. Accessed 20 May 2021
- @souyakuchan. <https://twitter.com/souyakuchan>. Accessed 29 Mar 2021
- Statistics of souyakuchan magic. https://kubor.github.io/posts/souyakuchan_stats.html. Accessed 11 Jan 2021
- Twitter-integrated 3D molecule viewer. <https://chemical.space/viewer>. Accessed 28 Feb 2021
- @retrosynthchan. <https://twitter.com/retrosynthchan>. Accessed 29 Mar 2021
- Drug Discovery (dry) Advent Calendar 2020. <https://adventar.org/calendars/5119>. Accessed 29 Mar 2021
- Genheden S, Thakkar A, Chadimová V, Reymond J-L, Engkvist O, Bjerrum E (2020) AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. *J Cheminform* 12(1):1–9. <https://doi.org/10.1186/s13321-020-00472-1>
- Weininger D (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J Chem Inf Comput Sci* 28(1):31–36. <https://doi.org/10.1021/ci00057a005>
- RDKit: open-source cheminformatics. <http://www.rdkit.org>
- POST statuses/filter. <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/filter-realtime/api-reference/post-statuses-filter>. Accessed 29 Mar 2021
- Tweepy: an easy-to-use Python library for accessing the Twitter API. <https://www.tweepy.org/>. Accessed 29 Mar 2021
- POST statuses/update. <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/post-and-engage/api-reference/post-statuses-update>. Accessed 29 Mar 2021
- Pillow (PIL Fork) 8.1.2 documentation. <https://pillow.readthedocs.io/>. Accessed 29 Mar 2021
- Developer Document for Twitter Player Card. <https://developer.twitter.com/en/docs/twitter-for-websites/cards/overview/player-card>. Accessed 28 Feb 2021
- Sehnal D, Deshpande M, Vařeková RS, Mir S, Berka K, Midlik A, Pravda L, Velankar S, Koča J (2017) LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat Methods* 14(12):1121–1122. <https://doi.org/10.1038/nmeth.4499>
- Murray-Rust P, Rzepa HS (1999) Chemical markup, XML, and the world-wide web. 1. Basic principles. *J Chem Inf Comput Sci* 39(6):928–942. <https://doi.org/10.1021/ci990052b>
- Bienfait B, Ertl P (2013) JSME: a free molecule editor in JavaScript. *J Cheminform* 5:24. <https://doi.org/10.1186/1758-2946-5-24>
- Csizmadia P (1999) MarvinSketch and MarvinView: molecule applets for the world wide web. In: Proceedings of the 3rd international electronic conference on synthetic organic chemistry. MDPI, Basel, Switzerland. <https://doi.org/10.3390/ecsoc-3-01775>
- Jiang C, Jin X, Dong Y, Chen M (2016) Kekule.js: an open source JavaScript cheminformatics toolkit. *J Chem Inf Model* 56(6):1132–1138. <https://doi.org/10.1021/acs.jcim.6b00167>
- rdkit/RDKit.js: JS wrappers around parts of the RDKit. <https://github.com/rdkit/RDKit.js>. Accessed 12 June 2021
- Dalby A, Nourse JG, Hounshell WD, Gushurst AKI, Grier DL, Leland BA, Laufer J (1992) Description of several chemical structure file formats used by computer programs developed at molecular design limited. *J Chem Inf Comput Sci* 32(3):244–255. <https://doi.org/10.1021/ci00007a012>
- pako: zlib port to javascript, very fast! <http://nodeca.github.io/pako/>. Accessed 29 Mar 2021
- brianloveswords/base64url: converting to, and from, base64url. <https://github.com/brianloveswords/base64url>. Accessed 29 Mar 2021
- Molecular sanitization—the RDKit book: the RDKit 2021.03.1 documentation. https://www.rdkit.org/docs/RDKit_Book.html. Accessed 29 Mar 2021
- The COVID Moonshot Consortium (2020) COVID Moonshot: open science discovery of SARS-CoV-2 main protease inhibitors by combining crowdsourcing, high-throughput experiments, computational simulations, and machine learning. *bioRxiv*. <https://doi.org/10.1101/2020.10.29.339317>
- Chodera J, Lee AA, London N, von Delft F (2020) Crowdsourcing drug discovery for pandemics. *Nat Chem* 12(7):581. <https://doi.org/10.1038/s41557-020-0496-2>
- PostEra: COVID Moonshot: an international effort to DISCOVER A COVID ANTIVIRAL. <https://covid.postera.ai/covid>. Accessed 29 Mar 2021

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

