



**HAL**  
open science

## The rise of preprints in chemistry

François-Xavier Coudert

► **To cite this version:**

François-Xavier Coudert. The rise of preprints in chemistry. *Nature Chemistry*, 2020, 12 (6), pp.499-502. 10.1038/s41557-020-0477-5 . hal-02713264

**HAL Id: hal-02713264**

**<https://hal.science/hal-02713264>**

Submitted on 1 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## The rise of preprints in chemistry

*Chemistry is now starting to embrace preprints, with more and more researchers in chemical and materials sciences posting their manuscripts online prior to peer review. Preprints can speed up the dissemination of scientific results and lead to more informal exchanges between researchers, hopefully accelerating the pace of research as a whole.*

François-Xavier Coudert

A preprint is a scientific article made available online to the research community by its authors that has not yet been formally peer reviewed (Fig. 1). Most often it is the version of an article that has just been (or soon will be) submitted to a traditional journal, but it can also sometimes be a working document that its authors wish to share publicly. Preprints allow researchers, amongst other things, to disseminate their latest advances in research more quickly, to stake a claim of priority on the research, and to receive feedback from other researchers in the wider community.

The online pre-publication model was invented by theoretical physicists, with the arXiv preprint server<sup>1</sup> that opened in 1991, hosted at the time by the Los Alamos National Laboratory. The scope of the server was later expanded to include contributions from other fields including astronomy, mathematics, computer science and quantitative biology to name just a few. In January 2015, arXiv passed the significant milestone of hosting one million articles and continues to grow, with 155,866 new submissions in 2019 (up 11% from the previous year). Its success has led to other discipline-centric preprint servers — including bioRxiv<sup>2</sup>, which was launched in 2013 for biology and has published 79,570 preprints since then.

The chemistry community came relatively late to preprinting, but the chemistry-focused chemRxiv<sup>3</sup> followed in August 2017 and currently hosts 4,391 preprints — a similar adoption rate, for its second anniversary, as was seen at bioRxiv. Additionally, there are other 'open archives' that accept preprints from any discipline — such as OSF Preprints<sup>4</sup> or HAL<sup>5</sup> — as well as institutional repositories. The recent upward trend<sup>6</sup> in preprinting across all fields of research comes at a time when the academic publishing system is witnessing a rapid acceleration towards more open access when it comes to scientific publishing — and the promotion of open science more generally.

**Faster dissemination of ideas**

In 2019 a survey of bioRxiv users<sup>7</sup> focused on researchers' motivations for posting preprints and the perceived benefits of doing so. One of the main reasons for preprinting research cited by those surveyed is the rapid dissemination of scientific knowledge to a (potentially) large audience: not only do preprints signal what work is coming out of a particular lab, but they also enable researchers to control when it becomes available. In the context of fast electronic communication being the norm in the modern world — combined with advances in research that can be really quite rapid — it is not surprising that editorial processes that can take several months are considered slow in comparison; colleagues complaining at the coffee machine or on social media about the insufficient pace of peer review is commonplace.

Just like sharing your latest results at a conference, uploading a preprint enables you to share and discuss a draft of your cutting-edge research. Moreover, informal feedback from those who read the preprint can often improve the quality of the final article: 71% of bioRxiv users say they received feedback on their preprints. Although some preprint servers have dedicated comment sections, this feedback most often takes place through other channels including social media, e-mail and discussion at conferences. For example, preprints are regularly publicized on Twitter by the @chemrxiv account, often leading to further discussion on that platform.

Preprinting also has other benefits. One that is frequently cited by researchers<sup>8</sup> is that uploading a preprint also makes it possible to claim priority on new ideas and developments — in a public and indisputable way — without depending on acceptance by the journal to which it will be submitted. This was already cited as one of the motivating factors back when arXiv was launched: some researchers at the time were concerned that unethical journal reviewers would try to delay publication with the intention of publishing the ideas or findings that they were meant to be assessing under their own name<sup>9</sup>. Although there is often little more than anecdotal evidence for such behaviour, preprinting the submitted version of a paper establishes a clear picture of the state of a piece of work at a specific date. More generally speaking, the use of preprints is a good way to record a timeline of ideas and knowledge as well as the evolution of papers that may change (and hopefully be improved) as a result of the peer-review process: in doing so, the practice of preprinting aligns well with the recent push by some journals to share the content of the peer-review reports and authors' responses to them. This reflects the nonlinear nature of research, where hypotheses are proposed, discussed, proven or refuted over time.

### **Increased visibility**

Several bibliometric studies have shown that preprints also increase the visibility of the work being done<sup>10</sup>, by combining two distinct advantages: they are open access and they appear online earlier than

the final peer-reviewed publication. This typically translates into more views and higher impact than non-preprinted articles in the same field<sup>11,12</sup>: preprinted articles typically have better online metrics, attention score, and number of citations.<sup>13</sup> Preprints for all servers are indexed by Google Scholar, and those deposited on chemRxiv are indexed in CAS and CrossRef (and Scopus is in the pipeline), so their discoverability is essentially the same as peer-reviewed papers. Articles can be cited in their preprint form: many servers now provide a DOI as a unique identifier for each preprint, making it easy to include them in reference lists. Citations specifically to the preprint version of an article are added to those of the final version of a paper that ultimately appears in a journal — an important factor at a time when bibliometrics play a significant role in the evaluation of research and researchers alike.

The use of preprints should be particularly attractive to early-career researchers, who typically have few published articles associated with their independent careers. It makes it possible to share articles that are in the process of being submitted in a more concrete fashion than the dreaded 'submitted for publication' phrase often found on CVs. More and more funding agencies accept (and even recommend) references to preprints in grant applications, as evidence of the applicant's previous work<sup>14</sup>: these include, amongst others, the European Research Council, US National Institutes of Health and the UK's Wellcome Trust. Hiring committees can also look favourably on this practice.

### **A step toward open science**

The use of preprints also ensures a wider circulation of the article's content — especially if the final published version appears in a subscription-based journal — because the preprint remains freely accessible after formal publication. Preprints can also be linked to the final published version of the article and can even be updated to the latest 'author accepted' version if the policy of the journal in question allows it. The dissemination of research results to the widest possible audience is considered to be a moral obligation of academic research<sup>15</sup>, especially when the majority of published research is funded from public monies<sup>16</sup>. Although a transition of the publishing system is under way, with a considerable push toward open access, preprints represent a cost-free way to achieve open access: it is a form of self-archiving by the authors, often referred to as 'green open access'. Preprints therefore provide a guarantee of access to those beyond well-funded research institutions, such as NGOs, patient organizations, journalists, independent researchers, researchers from developing countries and educators at non-research institutions. They also ensure the long-term sustainability of access to such files: preprint servers have an archiving and mirror system in place — and in many cases institutional guarantees. This is a marked difference with some commercial article-sharing platforms which sometimes require readers to create an account (so the content is not open access) and offer no long-term guarantee for the content that they host.

Moreover, the rapid sharing of research as it happens is in line with the idea of open science: preprints can help shape the future research in the field by making ideas spread faster. They contribute by filling a role that was traditionally served by oral and poster communications at conferences and workshops. Although conferences have increased in both size and number, preprints share none of the ecological, financial or personal costs of travel associated with those — and they also scale better by being available to all. This is particularly important with the current wave of travel restrictions related to the COVID-19 outbreak. Finally, it should also be noted that preprint servers also allow data associated with the research to be uploaded alongside the article itself, just as regular journals do. This is not limited to traditional supporting materials in PDF format, but also datasets of experimental and computational data, in machine-readable formats, enabling future researchers to build upon the work.

There are, of course, also some risks associated with the use of preprints. An important one is a potential blurring of the lines between peer-reviewed articles and preprints, meaning that preprints could create confusion in the eyes of journalists and the general public<sup>17</sup>. In particular, in topics with public-health implications it is important that preprints, if they are covered in the media, are not overblown — especially if, after review, the science turns out to be flawed in some way. Here it is important to note that, unlike published papers, preprints are typically viewed as part of the ongoing scientific discussion taking place in the community: since there is no formal peer review or quality control (beyond some basic moderation), there is no formal process for retraction or expression of concern on preprints posted online. However, we know that even peer-reviewed publications can later turn out to be incorrect, so perhaps this is an opportunity for a larger debate about the scientific method and how research works.

The COVID-19 outbreak and ensuing public health crisis is quite revealing of both the importance and limitations of preprint servers. While scientific journals provide accelerated peer review for work related to the current crisis, immediate and open access to the latest research data truly accelerates the pace of research, with analysis and critique often provided mere hours after preprints are posted. At the same time, the multitude of studies on a hot topic of such importance to public health has led preprint servers to add banners clearly reminding readers that, such as in the case of medRxiv, 'these are preliminary reports that have not been peer-reviewed. They should not be regarded as conclusive, guide clinical practice/health-related behavior, or be reported in news media as established information.'

### **Preprints in chemistry**

As noted earlier in this article, the chemistry community as a whole came relatively late to preprinting. In some subdisciplines, researchers have been uploading preprints for a while now, especially in areas

such as physical chemistry, computational chemistry and, more recently, biochemistry, but this did not become a common occurrence in our field until very recently<sup>18</sup>. In 2017, not one but two preprint servers for the chemical sciences were launched: chemRxiv, backed by a consortium of the American Chemical Society, the Royal Society of Chemistry, the German GDCh, the Chinese Chemical Society and the Chemical Society of Japan; and ChemRN, launched by Elsevier after its acquisition of SSRN (a preprint archive that previously focused on social sciences and law). These servers have met with success and the number of preprints in chemistry is now rapidly increasing. ChemRxiv published 2,314 new preprints in 2019 — compared to 1,031 in 2018 — and has seen more than 3 million article reads in the last year. This is still a drop in the ocean when it comes to chemistry publishing, with roughly 500,000 academic papers in chemistry and chemistry-related fields appearing each year. Nevertheless, the acceptance and adoption of the practice of preprinting is rapidly growing.

Parallel to this uptake, chemistry journals have clarified or revisited their positions regarding preprints. While some editors previously considered that uploading preprints compromised the novelty of the findings reported in a submitted manuscript, preprints are now universally welcomed by chemistry journals, from all different communities — and the vast majority of journals have a well-documented preprint policy, which can be found with online tools such as SHERPA/RoMEO<sup>19</sup>. The distribution of preprints among the different sub-fields of chemistry is currently biased towards fields that already practiced preprinting before the launch of the chemistry-specific servers, such as theoretical chemistry and materials science (Fig. 2). However, there is now a rapid rise in the areas of organic chemistry and chemical biology.

Preprints have made significant inroads into chemistry, but worries are occasionally expressed suggesting that the act of preprinting an article will, despite the assurances on journal websites, prejudice its chances of appearing in what are considered by some researchers to be high-profile venues for publication. The counter-argument to this is to highlight examples of published research papers that were first shared as preprints. There are many examples, but to select just a few, you can look to the 16-step synthesis of isoryanodane diterpene (+)-perseanol by Han and co-workers<sup>20,21</sup>, the CryoEM methodological paper by Jones and colleagues<sup>22,23</sup>, or the study of modular radical cross-coupling with sulfones by the Baran group and collaborators<sup>24,25</sup>.

The field of chemistry has seen a rapid increase in the adoption of preprints in the past couple of years, linked to the development of community preprint servers and the shift to preprint-friendly editorial policies by the journals in the field. Preprints, coupled with comment platforms and social media, allow for quick and broad dissemination of a research team's latest results as well as feedback from the community and the overall process replicates some aspects typically associated with conference talks, posters and even informal social interactions such as coffee breaks. They are an established part of the

scientific conversation in many disciplines and there is no reason why chemists should not take advantage of the benefits they offer, particularly when it comes to giving credit to early career researchers. The way in which scientific results are disseminated has changed significantly with the advent of the internet and preprints represent an important step in the evolution of scholarly communication.

François-Xavier Coudert

Chimie ParisTech, PSL University, CNRS, Institut de Recherche de Chimie Paris, 75005 Paris, France.

### Competing interests

François-Xavier Coudert is a member of ChemRxiv Scientific Advisory Board, and Editorial Board Member for *Communications Chemistry*.

### References

1. Available online at <https://arxiv.org>
2. Available online at <https://www.biorxiv.org>
3. Available online at <https://chemrxiv.org>
4. Available online at <https://osf.io/preprints/>
5. Available online at <https://hal.archives-ouvertes.fr>
6. Fry, J., Spezi, V., Probets, S. & Creaser, C. *J. Assoc. Info. Sci. Tech.* **67**, 2710–2724 (2016).
7. Sever, R. et al. *bioRxiv* <https://doi.org/10.1101/833400> (2019).
8. Fowler, K. K. *Issues in Science and Technology Librarianship* <https://doi.org/10.5062/F4QN64NM> (2011).
9. Vence, T. Q&A: 1 Million Preprints and Counting. *TheScientist* December 29, 2014. (Available at: <https://www.the-scientist.com/daily-news/qa-1-million-preprints-and-counting-36168>).
10. Shuai, X. Pepe, A., Bollen, J. & Ouzounis, C. A. *PLOS One*, **7**, e47523 (2012).
11. Youngen, G. K. *Coll. Res. Libr.* **59**, 448–456 (1998).
12. Craig, I. D., Plume, A. M., McVeigh, M. E., Pringle, J. & Amin, M. *J. Informetr.* **1**, 239–248 (2007).
13. Serghiou, S. & Ioannidis, J. P. A. *JAMA* **319**, 402–404 (2018).
14. Matthews, D. New boost for preprints after acceptance by ERC. *Times Higher Education* August 9, 2018. (Available at <https://www.timeshighereducation.com/news/new-boost-preprints-after-acceptance-erc>).
15. Pearn, J. *BMJ* **310**, 1313–1315 (1995).

16. Mervis, J. *Science* <https://doi.org/10.1126/science.aal0890> (2017).
17. Sheldon, T. *Nature* **559**, 445–445 (2018).
18. Velden, T. & Lagoze, C. *Nature Chem.* **1**, 673–678 (2009).
19. Available online at <http://sherpa.ac.uk/romeo/index.php>
20. Han, A., Tao, Y. & Reisman, S. E. *Nature* **573**, 563–567 (2019).
21. Available on chemRxiv at <https://doi.org/10.26434/chemrxiv.8100152>
22. Jones, G. J. et al. *ACS Cent. Sci.* **4**, 1587–1592 (2018).
23. Available on chemRxiv at <https://doi.org/10.26434/chemrxiv.7215332>
24. Merchant, R. R. et al. *Science* **360**, 75–80 (2018).
25. Available on chemRxiv at <https://doi.org/10.26434/chemrxiv.5715106>

**Figure 1 | How preprints fit into the ecosystem of academic publishing.** The process in the top part of the figure shows the traditional route to publication in which a manuscript is submitted to a journal and then undergoes editorial and peer evaluation before finally being accepted for publication (and not necessarily in the first journal it was submitted to). The lower-half shows how preprints fit into this model, comparing the availability of preprints and journal articles to be read (and cited). Preprints are available immediately and any community feedback can be used to help in revising the manuscript during the formal evaluation at a journal.



**Figure 2 | Distribution of broad subject areas of 4,391 preprints posted on chemRxiv as of April 2020.** There is a wide spread of subject areas, but the largest one is in the subdiscipline of chemistry (theory and computation) that has traditionally embraced preprinting and open-science principles more so than other areas.