

EDITORIAL

Open Access



Are new ideas harder to find? A note on incremental research and *Journal of Cheminformatics*' Scientific Contribution Statement

Barbara Zdrzil^{1*} , Rajarshi Guha² , Karina Martinez-Mayorga³  and Nina Jeliaskova⁴

In the field of cheminformatics, technological advancements in recent times include, e.g., the way chemical information is being represented for large scale screening and *de novo* drug design. Especially, chemical language models originating from natural language processing offer new opportunities for molecular design [1].

However, for science in general and compared to past decades, recent paucity of transformative ideas has been noticed [2]. While there are many explanations for observed technological stagnation, in pharmaceutical R&D, a productivity crisis was already noted ~20 years ago [3, 4]. An often stated scientific/technological reason for stagnation in pharmaceutical R&D, is the “low hanging fruit” problem. That is, the easier-to-tackle problems have been solved already and that what remains are the more complex and more challenging problems (diseases) [5]. Other possible explanations for declining research productivity might be a shift to more “defensive R&D”, which could be a direct consequence of R&D resources being diverted away from risk-taking by investors,

managers, and entrepreneurs to instead fulfil regulatory requirements. Instead of fueling innovation, monetary resources are used to keep “old” products on the market [5].

At the same time, we do observe a trend that more and more papers are being published in scientific journals or on preprint servers [6]. In line with this observation, also more data, methods, and models are being made available in the public domain (through publications and/or platforms such as Zenodo [7], GitHub [8], and Hugging Face [9]). As an effect, researchers are often facing a situation of information-overload with the luxurious problem of filtering out the real innovative contributions, that aren't just incremental improvements of existing ones.

From a publisher's perspective, every research paper should be regarded as an attempt to contribute new ideas and/or refine old ones. In Cheminformatics, we have observed a few phases of new methodological developments/inventions with consequent iterations of incremental improvements. Examples include (but are not limited to) molecular representation [10], descriptors for QSPR modeling/ML [11, 12], molecular docking algorithms [13], or more recently the development and refinement of generative AI algorithms [14, 15].

The Editors of the *Journal of Cheminformatics* do not judge articles based purely on scientific novelty. Rather, we consider aspects such as utility and availability, and the contribution itself, along with notions of novelty.

The Scientific Contribution Statement is our attempt to give space to a declaration by authors regarding the

*Correspondence:

Barbara Zdrzil
bzdrzil@ebi.ac.uk

¹ European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, UK

² Vertex Pharmaceuticals, 50 Northern Ave, 02210 Boston, MA, USA

³ Institute of Chemistry, National Autonomous University of Mexico, Campus Merida, Merida-Tetiz Highway, Km. 4.5, Ucu, Yucatan, Mexico

⁴ Ideacon Ltd, 1000 Sofia, Bulgaria



© The Author(s) 2024. **Open Access** 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.

contributions made in their research. The authors should use a maximum of three sentences to specifically highlight the scientific contributions that advance the field and what differentiates their contribution from prior work on this topic (<https://jcheminf.biomedcentral.com/submission-guidelines/preparing-your-manuscript/research>). It should be regarded by authors as an opportunity to highlight their scientific contribution(s) rather than as a burden or additional request by the Editors of *J. Cheminform.* Such declaration(s) about contributions and novelty have always been part of the scientific publication process—albeit in a more convoluted or scattered way, as there usually isn't a specific section in a paper dedicated to such declarations. We therefore started to make this vital information more accessible by assigning it a fixed section (namely the Abstract of a paper).

We hope that this amendment will not only help us as Editors when assessing a paper submitted for consideration, but equally the members of our scientific community—reviewers and readers.

Author contributions

BZ drafted the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests

The authors declare that they have no competing interests.

Published online: 15 January 2024

References

1. Grisoni F (2023) Chemical language models for de novo drug design: challenges and opportunities. *Curr Opin Struct Biol* 79:102527. <https://doi.org/10.1016/j.sbi.2023.102527>
2. Bhaskar M (2021) Human frontiers: the future of big ideas in an age of small thinking
3. Pammolli F, Magazzini L, Riccaboni M (2011) The productivity crisis in pharmaceutical R&D. *Nat Rev Drug Discov* 10:428–438. <https://doi.org/10.1038/nrd3405>
4. Laermann-Nguyen U, Backfisch M (2021) Innovation crisis in the pharmaceutical industry? A survey. *SN Bus Econ* 1:164. <https://doi.org/10.1007/s43546-021-00163-5>
5. Cockburn IM (2006) Is the pharmaceutical industry in a productivity crisis? *Innov Policy Econ* 7:1–32. <https://doi.org/10.1086/ipe.7.25056188>
6. Rawat S, Meena S (2014) Publish or perish: where are we heading? *J Res Med Sci* 19:87–89
7. <https://zenodo.org/>
8. <https://github.com/>
9. <https://huggingface.co/>
10. Wigh DS, Goodman JM, Lapkin AA (2022) A review of molecular representation in the age of machine learning. *WIREs Comput Mol Sci* 12:e1603. <https://doi.org/10.1002/wcms.1603>
11. Balaban AT (2009) Drug Design, Molecular descriptors in. In: Meyers RA (ed) *Encyclopedia of complexity and Systems Science*. Springer, New York, NY, pp 2196–2215
12. Sahoo S, Adhikari C, Kuanar M, Mishra BK (2016) A short review of the generation of Molecular descriptors and their applications in quantitative structure Property/Activity relationships. *Curr Comput Aided Drug Des* 12:181–205. <https://doi.org/10.2174/1573409912666160525112114>
13. Torres PHM, Sodero ACR, Jofily P, Silva-Jr FP (2019) Key topics in molecular docking for drug design. *Int J Mol Sci* 20:4574. <https://doi.org/10.3390/ijms20184574>
14. Bandi A, Adapa PVSR, Kuchi YEVPK (2023) The Power of Generative AI: a review of requirements, models, input–output formats, evaluation Metrics, and challenges. *Future Internet* 15:260. <https://doi.org/10.3390/fi15080260>
15. Shokrollahi Y, Yarmohammadtoosky S, Nikahd MM et al (2023) A comprehensive review of generative AI in healthcare

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

