

a new expression product. Such an event could be of no consequence, or it could affect the safety of food derived from the animal. We can't know this if we don't look. Consumers expect the FDA to ensure their food is safe and our international trading partners expect that the FDA will have completed a scientific evaluation that determines that exported products are safe.

The FDA strongly supports the use of genome editing and other technologies to bring new, innovative products to market that can improve human and animal health, improve animal well-being, and enhance food production and quality. The agency is committed to doing so in an efficient, flexible, risk-based manner, based on sound science, that will ensure the products are safe for animals and people and that they achieve their claimed results. These potential results promise great benefits, including improvements to human health, such as reduced allergenicity and transmission of disease; to animal health, such as disease resistance and heat tolerance; to food production, such as enhanced nutrition and improved growth; and to human therapeutic production, such as new, more efficient production methods and more appropriate models of disease for use in research.

At this early stage, as genome-editing technology is continuing to develop and the science is evolving, bringing products with unknown risks to market without adequate oversight to ensure they are safe and that they produce the promised effects will undermine consumer confidence and, ultimately, set back the progress of the entire field. As this promising scientific area further progresses and researchers continue to optimize genome-editing technology, we will continue to identify products that are low risk. This will enable

the FDA to direct our regulatory activities away from those low-risk products and toward higher-risk products.

The FDA intends to prioritize use of its resources in a manner that is public health protective and risk based. It makes no sense to expend the FDA's limited resources taking action against low-risk products. But it also means that, at a time when the science is still evolving and a full understanding of risk is developing, the FDA will not shirk its responsibility to examine the data, identify the risks and determine whether they pose a real hazard to people or animals.

We recognize that there is tremendous excitement over quickly embracing and bringing to market the fruits of genome-editing technology, as well as the critical importance of adequately identifying potential risks, efficiently evaluating whether the risks do in fact exist, and determining whether the risks pose an actual safety hazard in a timely manner. It is the FDA's role to balance these competing imperatives. We want to support the timely development of beneficial products, but not at the expense of abdicating our critical role in protecting consumer and animal health. We have a public health obligation to protect consumer and animal health that we must balance with the potential of this innovative technology to enhance human and animal health and food production.

This balance is not new to the FDA. It is one that is inherent in the agency's role as public health regulator in a sector that, over time, employs new technologies and methods to produce advancements that no one could have envisioned at the agency's start. There are a host of innovations — nanotechnology, stem cell therapies, computer software, and so on — to which the FDA has applied its scientific expertise

to ensure the products produced using such novel methods are safe and effective. Just as we've integrated the regulation of other innovative products into the FDA's critical work, the FDA has integrated the regulation of intentional genomic alterations in animals into its veterinary regulatory program. The FDA stands ready to work with developers to bring forward promising new agricultural biotechnology products.

Toward this end, we have initiated a Veterinary Innovation Program to provide an 'all hands on deck' approach to help developers efficiently navigate the regulatory process, and we are already working with several developers as part of this new program. The FDA has the appropriate expertise and experience to regulate intentional genomic alterations in animals developed using novel techniques, such as genome editing, through a timely and efficient process, and we are committed to doing so in a manner that fosters innovation, promotes consumer confidence and protects the public health. □

Steven M. Solomon

Center for Veterinary Medicine, US Food and Drug Administration, Rockville, MD, USA.

e-mail: Steven.Solomon@fda.hhs.gov

Published online: 7 February 2020

<https://doi.org/10.1038/s41587-020-0413-7>

References

- Norris, A. et al. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-019-0394-6> (2020).
- Carlson, D. F. et al. *Nat. Biotechnol.* **34**, 479–481 (2016).
- Nagahata, H. *J. Vet. Med. Sci.* **66**, 1475–1482 (2004).
- Shuster, D. E., Kehrl, M. E. Jr., Ackermann, M. R. & Gilbert, R. O. *Proc. Natl Acad. Sci. USA* **89**, 9225–9229 (1992).
- Kumar, V. & Sharma, A. *Agric. Rev. (Karnal)* **30**, 293–300 (2009).
- Olson, E. N., Arnold, H.-H., Rigby, P. W. & Wold, B. J. *Cell* **85**, 1–4 (1996).
- Meier, I. D. et al. *FASEB J.* **24**, 1714–1724 (2010).

Assessing the impact of generative AI on medicinal chemistry

To the Editor — The profound challenges of drug discovery, coupled with the societal importance of the task, make it imperative that we investigate novel, creative methods that improve our abilities to design new medicines. In recent years, attempts at developing and deploying a wide range of computational methods to support drug discovery have accelerated—sometimes with extraordinary claims made about

their significance. Novel computational approaches require rigorous evaluation to determine their true utility in real-world drug discovery settings. Sadly, novel methods often are disclosed without sufficient documentation, making it difficult or impossible to carry out such an objective evaluation.

Over the past few years, interest has grown in the application of artificial

intelligence (AI) techniques to drug discovery^{1,2}. One active branch of AI that has been the focus of a tremendous amount of recent activity is the field of generative modeling^{3–9}. In this technique, a deep learning model is trained based on a corpus of existing molecules. The model typically 'encodes' a higher-dimensional representation, such as a SMILES (simplified molecular-input line-entry system)¹⁰, into

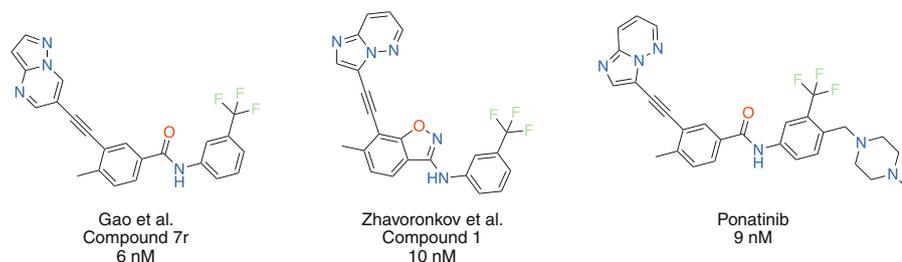


Fig. 1 | Comparison of compound 1 and two related inhibitors. The chemical structure of compound 1 from Zhavoronkov et al. compared with those of ponatinib, a marketed multi-kinase inhibitor, and a DD1 inhibitor reported by Gao et al.¹⁴ in 2013.

a lower-dimensional representation, often referred to as a latent space. This latent space can then be ‘decoded’ back to the higher-dimensional representation to create new molecules. The exploration of this latent space can be coupled with a predictive model with the aim of discovering novel, active molecules. In a sense, generative models can be seen as a variation on the de novo design¹¹ programs that were in vogue during the 1990s and early 2000s. As with de novo design, evaluating the significance of the output of these models is not straightforward. Although two groups have made initial efforts at developing methods for benchmarking generative models^{12,13}, evaluating the novelty, and ultimately the significance, of the molecules generated by these methods remains an open question. Such benchmarks provide a common ground for evaluation and comparison, but the ultimate value of generative models will be demonstrated through the synthesis and biological evaluation of the novel molecules they identify.

One of the first examples of the synthesis and testing of molecules derived from a generative model was reported in a 2018 paper by Merk and co-workers¹⁴. In this paper, the authors began by training a generative model on a set of >500,000 bioactive molecules from the ChEMBL database. They then fine-tuned the model to generate molecules that would be agonists of retinoid X receptors (RXRs) or peroxisome proliferator-activated receptors (PPARs). This fine-tuning process involved training based on a set of 25 fatty acid mimetics known to be agonists of PPAR or RXR. The molecules produced by the generative model were evaluated based on a quantitative structure–activity relationship (QSAR) model for PPAR and RXR activity. The authors then used the rankings from the QSAR model, along with a manual assessment of synthetic accessibility and chemical building block availability, to

select five molecules to be synthesized. Although the authors reported that the five molecules selected for synthesis were not found in databases of reported molecules, they did not report the structures of the 25 fatty acid mimetics used in the training process. The five selected molecules were then synthesized, and two were found to be PPAR agonists with half-maximal effective concentration (EC₅₀) values between 4 μM and 14 μM. Two additional compounds were dual PPAR and RXR inhibitors with EC₅₀ values between 60 nM and 13 μM. The fifth compound was reported as inactive.

Another, more recent example of the synthesis of a set of compounds based on a generative model comes from a paper by Zhavoronkov et al.¹⁵ published in the September issue of *Nature Biotechnology*. In this paper, the authors trained a generative model based on a large set of discoidin domain receptor family member 1 (DDR1) inhibitors extracted from the scientific and patent literature. Based on the output of the generative model, they synthesized six molecules. Of these six molecules, four had measurable biochemical activity, with the best, ‘compound 1’, having a 10 nM biochemical half-maximal inhibitory concentration (IC₅₀). Compound 1 was then tested in the U2OS bone osteosarcoma cell line and shown to have an IC₅₀ of 10.3 nM. In a subsequent mouse pharmacokinetic study, it was also shown to have reasonable bioavailability as well as a half-life of 3.5 h.

This result received quite a bit of notice in the scientific and popular press and was hailed as ‘revolutionary’ by several pundits, some of them with clear competing financial interests¹⁶. An investor in the company that published the paper went as far as to refer to this as “Pharma’s AlphaGo Moment,” referring to the recent case in which Google’s program AlphaGo¹⁷ was able to defeat grandmaster Fan Hui in the challenging strategy game known as Go. One fact that seems to have escaped most of these pundits is the striking similarity

between compound 1 in the Zhavoronkov paper and ponatinib (marketed as the drug Iclusig; Fig. 1).

Ponatinib was originally developed by ARIAD Pharmaceuticals as a BCR-ABL inhibitor¹⁸ for the treatment of acute lymphocytic leukemia and chronic myelogenous leukemia. Subsequent testing showed that the compound was a potent inhibitor of multiple tyrosine kinases¹⁹. This broad kinase profile is believed to be the cause of undesirable side effects which led the US Food and Drug Administration to assign ponatinib a black box warning in 2013. One of the many kinases inhibited by ponatinib is DDR1, the kinase used as the target in the paper by Zhavoronkov et al. Several papers, including those used to supply the corpus for the Zhavoronkov paper^{18,20}, list the DDR1 IC₅₀ of ponatinib at between 1 nM and 9 nM. Given this similarity to a marketed multi-kinase inhibitor, the cellular and pharmacokinetic profiles of compound 1 become considerably less surprising. It should also be noted that an analog of ponatinib that slightly modifies the imidazopyridazine moiety was published as a 6 nM inhibitor of DDR1 in a 2013 paper by Gao and co-workers¹⁸. This paper was also part of the corpus of DDR1 inhibitors used to train the Zhavoronkov model.

Given the multi-kinase activity of ponatinib, one has to question the selectivity of compound 1 reported by Zhavoronkov et al. In their paper, the authors claim that compound 1 is selective. This claim is supported by a selectivity screen of 44 kinases, which is reported in Supplementary Fig. 8 in their paper. Unfortunately, this selectivity screen does not contain any of the kinases that have been reported¹⁹ as targets for ponatinib with IC₅₀ values between 1.5 and 72.2 nM (such as fibroblast growth factor receptor (FGFR) 1, FGFR2, FGFR3, FGFR4, platelet-derived growth factor receptor α (PDGFRα), vascular endothelial growth factor receptor 2 (VEGFR2), c-SRC, c-KIT, FMS-like receptor tyrosine kinase-3 (FLT3), RET and ABL). Without testing of these known ponatinib off-targets, claims of selectivity are difficult to defend.

The similarity of molecules produced by the generative model to known active molecules in the training corpus raises several questions that apply more generally to other papers identifying inhibitors using generative models.

The first question is the level at which the training data used for a generative model should be presented. Although Zhavoronkov et al. do provide the references for the molecules that went into their training set, they have not made the complete training

corpus available for comparison. In addition, the authors do not show ponatinib or any of its analogs in the paper or in the supporting material. We strongly believe that for results from a generative model to be published, the full training set used to build the models should be made available in electronic form. In addition, the training set molecules most similar to the final molecule(s) should be shown as chemical structures in the paper.

Perhaps a more important question is the criteria by which molecules produced by generative models should be judged. Will the requirements for novelty, activity and breadth of structure–activity relationship be the same as those for teams of human chemists? One has to ask whether a paper reporting work in which a team of chemists substituted an isoxazole for an amide carbonyl to generate a compound that is roughly equipotent with published compounds would be reviewed, let alone published.

Another issue that must be considered when assessing the performance of a generative model is whether a simpler approach that does not employ AI could have produced the same molecules. For years, computational chemists have employed approaches that automatically insert isosteric replacements. In some cases, these replacements are based on medicinal chemistry precedent²¹, whereas in other cases, molecular fragments with similar shape and electrostatics are substituted²². Another common approach is to employ a ring-bracing²³ strategy to reduce a molecule's conformational flexibility. Although it may be difficult to perform a head-to-head comparison, the existence of these alternatives should be noted.

In 2010, Stahl and Bajorath²⁴ published a clear set of guidelines for computational papers that were to be submitted for publication in the *Journal of Medicinal Chemistry* (JMC). These guidelines, which were subsequently integrated into the journal's Instructions for Authors, were key to improving the quality of computational medicinal chemistry papers in JMC and ensuring that computational papers were reviewed in a consistent fashion. We have reached a point where scientific journals need to put similar guidelines in place for

generative models. This will enable the more rapid and systematic evaluation of such methods, which will benefit the entire community.

As such, we would suggest that journals that publish results of generative models and related studies join together to create a set of guidelines for the review of papers disclosing molecules generated by AI methods. Although the guidelines will have to be a result of input and discussion from the community, we would like to suggest three guidelines:

1. The active molecules used to train the generative model should be made available in electronic form. This availability of data will make it easy for most readers to perform substructure and similarity searches and compare the output molecules with the training set.
2. Papers reporting AI-generated molecules should contain a table showing the training-set molecule most similar to each generated molecule. There are many ways to assess molecular similarity, and although this table is no substitute for the structure disclosure (point 1), casual readers will be able to quickly assess the significance of the reported molecules.
3. Journals should use the same criteria for assessing the novelty of AI-generated molecules that are used to assess molecules generated by a team of medicinal chemists. We hope to see a day when AI-generated molecules are indistinguishable from those arising from the creativity of medicinal chemists.

Although it would be beneficial to be able to compare the results of generative models with those of isosteric replacement and other de novo design methods, we do not see a practical way to implement this. De novo design software still tends to be somewhat of a bespoke solution, and these methods are not widely available.

AI methods have moved beyond the specialist realm that they have occupied for the past 30 years and are poised to become an integral part of the process of scientific discovery, and more specifically drug discovery. However, for these methods to

become mainstream, it is essential that we clearly separate hype from genuine utility. We need to be transparent about how the methods are being used and provide a clear context around the discoveries that are being made. The development of guidelines for publication of the results of generative models will enable reviewers and readers to more accurately assess and appreciate this new and rapidly evolving area of science. □

W. Patrick Walters[✉] and Mark Murcko
Relay Therapeutics, Cambridge, MA, USA.

*e-mail: pwalters@relaytx.com

Published online: 30 January 2020
<https://doi.org/10.1038/s41587-020-0418-2>

References

1. Griffen, E. J., Dosseter, A. G., Leach, A. G. & Montague, S. *Drug Discov. Today* **18**, 725–731 (2018).
2. Vamathevan, J. et al. *Nat. Rev. Drug Discov.* **18**, 463–477 (2019).
3. Putin, E. et al. *Mol. Pharm.* **15**, 4386–4397 (2018).
4. Segler, M. H. S., Kogej, T., Tyrchan, C. & Waller, M. P. *ACS Cent. Sci.* **4**, 120–131 (2018).
5. Jørgensen, P. B., Schmidt, M. N. & Winther, O. *Mol. Inform.* **37**, 1700133 (2018).
6. Schneider, G. *Mol. Inform.* **37**, 1880131 (2018).
7. Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A. & Zhavoronkov, A. *Mol. Pharm.* **14**, 3098–3104 (2017).
8. Dimitrov, T., Kreisbeck, C., Becker, J. S., Aspuru-Guzik, A. & Saikin, S. K. *ACS Appl. Mater. Interfaces* **11**, 24825–24836 (2019).
9. Gómez-Bombarelli, R. et al. *ACS Cent. Sci.* **4**, 268–276 (2018).
10. Weininger, D. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
11. Schneider, G. & Fechner, U. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).
12. Brown, N., Fiscato, M., Segler, M. H. S. & Vaucher, A. C. *J. Chem. Inf. Model.* **59**, 1096–1108 (2019).
13. Polykovskiy, D. et al. Preprint at <https://arxiv.org/abs/1811.12823v3> (2018).
14. Merk, D., Friedrich, L., Grisoni, F. & Schneider, G. *Mol. Inform.* **37**, 1700153 (2018).
15. Zhavoronkov, A. et al. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
16. Colangelo, M. *LinkedIn* <https://www.linkedin.com/pulse/pharmas-alpha-go-moment-first-time-ai-has-designed-new-colangelo/> (3 September 2019).
17. Silver, D. et al. *Nature* **529**, 484–489 (2016).
18. Gao, M. et al. *J. Med. Chem.* **56**, 3281–3295 (2013).
19. Tan, F. H., Putoczki, T. L., Stylli, S. S. & Luwor, R. B. *Oncotargets Ther.* **12**, 635–645 (2019).
20. Canning, P. et al. *J. Mol. Biol.* **426**, 2457–2470 (2014).
21. Stewart, K. D., Shiroda, M. & James, C. A. *Bioorg. Med. Chem.* **14**, 7011–7022 (2006).
22. Hessler, G. & Baringhaus, K. *Drug Discov. Today Technol.* **7**, e263–e269 (2010).
23. Leach, A. R. & Lewis, R. A. *J. Comput. Chem.* **15**, 233–240 (1994).
24. Stahl, M. & Bajorath, J. *J. Med. Chem.* **54**, 1–2 (2011).

Competing interests

W.P.W. is employed by, and M.M. is a member of the Board of Directors of, Relay Therapeutics.