Research briefing

'Explainable' AI identifies a new class of antibiotics

An artificial-intelligence graph neural network was trained on experimental data and used to identify chemical substructures that underlie selective antibiotic activity in more than 12 million compounds. This led to the discovery of a class of antibiotics with *in vitro* and *in vivo* activity against Gram-positive bacteria, including *Staphylococcus aureus*.

The mission

More and more disease-causing bacteria are becoming resistant to antibiotics. The number of deaths caused by infections with antibiotic-resistant microorganisms is projected to reach 10 million a year globally by 2050 (see go.nature.com/41uemys). One way to address this crisis of microbial resistance involves the discovery of new kinds of antibiotics. Although a few promising candidates have been identified in the past decade, other approaches to antibiotic discovery are urgently needed because conventional pipelines can no longer efficiently produce antibiotic[1].

Artificial intelligence (AI) approaches can increase the efficiency of discovering antibiotic candidates[2]. For instance, a type of model called a graph neural network can identify the arrangements of atoms and bonds in chemical structures that make it more likely that a compound has antibacterial activity[3]. In computational science, graphs represent data comprising vertices and edges, and graph neural networks are deep-learning models — algorithms with many data-processing steps — that can infer information, such as antibacterial activity, from chemical structures represented as graphs. We aimed to develop graph neural networks that could predict new structural classes of antibiotics.

The solution

Our approach integrates experimental and computational data. We experimentally evaluated the ability of 39,312 structurally diverse small-molecule compounds (with molecular masses of 40–4,200 daltons) to stop the growth of *Staphylococcus aureus* in a laboratory culture, as well as their tendency to kill three human cell types (liver carcinoma cells, primary skeletal muscle cells and lung fibroblasts). We then trained graph neural networks on each of the data sets and used the resulting models to predict the antibiotic activity of 12,076,365 compounds and their toxicity to human cells (Fig. 1a). We filtered the compounds on the basis of high predicted antibiotic activity, low predicted cytotoxicity, high structural novelty and favourable medicinal chemistry properties. After filtering, 3,646 compounds remained.

AI models, including graph neural networks, are often considered 'black boxes', because their predictions cannot be readily explained or interpreted. To understand the chemical substructures associated with high prediction scores for antibiotic activity, we 'opened' the black box by using a search method to computationally identify the models' 'rationales'. Our search method repeatedly removes atoms and bonds, and scores the resulting substructures to determine whether they are responsible for a large fraction of the prediction score. Similar search methods have been used for other deep-learning models, such as AlphaGo[4]. Notably, the identified rationales represent conserved structural features associated with antibiotic activity, and they predict structural classes of antibiotics. We computed rationales for 380 filtered compounds and, after experimentally testing this subset to demonstrate that our approach substantially increases true discovery rates, we focused on one promising rationale – the one shared by compounds **1** and **2** (Fig. 1b). Detailed studies of these compounds showed that they are effective against multidrug-resistant bacteria, including methicillin-resistant *S. aureus* (MRSA, a bacterium that causes conditions such as abscesses and sepsis). The compounds prevent antimicrobial resistance by disrupting the pH gradient across the bacterial membrane. Notably, they have favourable medicinal properties. Finally, we found that compound **1** effectively decreased MRSA levels in two mouse models *in vivo*.

The implications

Our approach enables the AI-guided discovery of structural classes of antibiotics by providing an efficient and powerful way of mining the chemical space – the set of all possible chemical compounds with determined properties. The rationale search, cytotoxicity models and chemical filtering steps might help with other discovery efforts, such as those for antiviral, anticancer and senolytic drugs. For computational science, our work shows that deep-learning models used for drug discovery can be made explainable, providing chemical insights into what has been learnt by often-undecipherable algorithms.

Our approach is limited by the specificity of the biological activity being predicted and the use of chemical structures as inputs. Predicting antibacterial activity against several pathogens and in various biological contexts, such as in the presence of serum proteins, remains important for assessing the compounds' clinical potential. Our models are also mechanism agnostic and do not use other available information, such as data on molecular-binding targets.

Felix Wong and **James J. Collins** are at the Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. **F.W.** is also at Integrated Biosciences, San Carlos, California, USA.

EXPERT OPINION

> This paper presents a coherent deep-learning system that may well improve the odds of discovering new antibacterial classes in large compound libraries compared with random screening. The findings will undoubtedly be of interest to researchers in the area of drug discovery." (CC BY 4.0)

Lynn Silver is at LL Silver Consulting, Springfield, New Jersey, USA.

REFERENCES

1. Lewis, K. *Cell* **181**, 29–45 (2020).

2. Wong, F., de la Fuente-Nunez, C. & Collins, J. J. *Science* **381**, 164–170 (2023).

3. Stokes, J. M. *et al. Cell* **180**, 688–702 (2020).

4. Silver, D. *et al. Nature* **550**, 354–359 (2017).
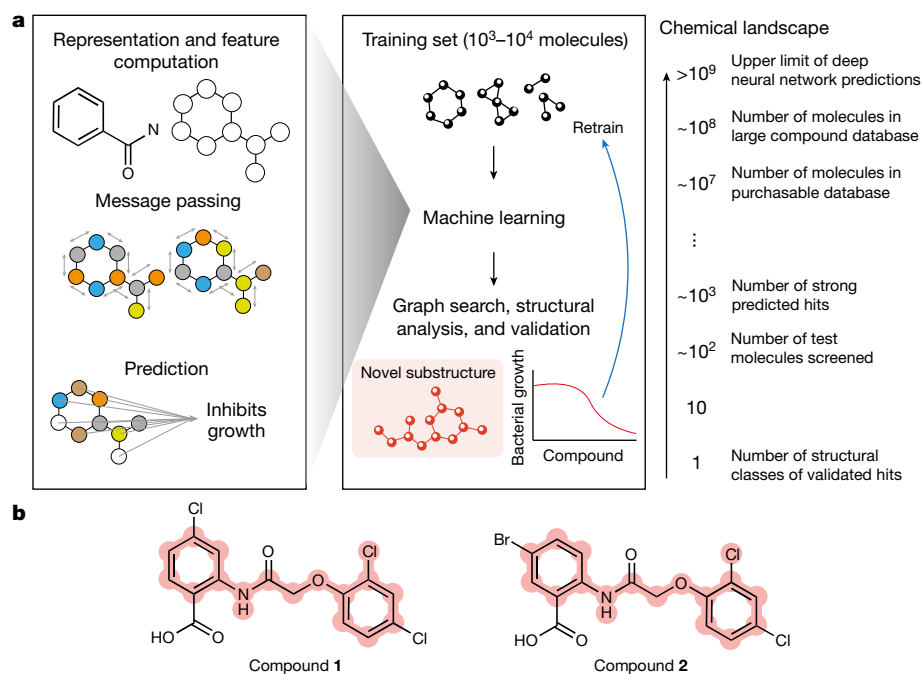
FIGURE

Figure 1 | A platform for identifying structural classes of antibiotics uses explainable deep learning. a, Schematic of the approach. We screened 39,312 compounds for antibacterial activity against *Staphylococcus aureus* RN4220. The data were used to train graph neural networks, which were used to make predictions of antibacterial activity for 12,076,365 molecules. By computing 'rationales' – chemical substructures that had high antibacterial-activity scores – our approach identifies predicted structural classes. **b**, Chemical structures of compounds **1** and **2**, which have a shared rationale (red substructure). These compounds have antibacterial activity against *S. aureus in vitro* and *in vivo*.

BEHIND THE PAPER

As a proof of concept, our laboratory previously used graph neural networks trained on growth-inhibition data from *Escherichia coli* to discover the inhibitor halicin[3]. Here, we have substantially expanded on our approach. First, we developed a method of 'opening the black box' to explain deep-learning model predictions, which helps with the identification of possible structural classes. Second, we established deep-learning models to predict cytotoxicity and then tested these experimentally to identify selective antibacterial compounds. Finally, we incorporated medicinal-chemistry filters to reflect drug-discovery knowledge.

Our approach resulted in the discovery of two promising antibiotic candidates, compounds **1** and **2**. Of those, compound **1** was effective in *in vivo* models of mice infected with MRSA, including a thigh-infection model that is difficult to treat. Developing our platform required years of persistent efforts by a diverse team of specialists.

F.W.

FROM THE EDITOR

Given how difficult it is to identify new classes of antibacterial leads, and the promise of AI in tackling this challenge, this work represents a potentially important blueprint for the rapid exploration of chemical space. The study will enable the prioritization of new and non-toxic compounds for further development.

Editorial team, *Nature*