Essay: Using Machine Learning for Antibiotic Discovery

Cesar de la Fuente-Nunez[1,2,3,4,*] and James J. Collins[5,6,7,†]

[1]Machine Biology Group, Departments of Psychiatry and Microbiology, Institute for Biomedical Informatics, Institute for Translational Medicine and Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[2]Departments of Bioengineering and Chemical and Biomolecular Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[3]Department of Chemistry, School of Arts and Sciences, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[4]Penn Institute for Computational Science, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[5]Institute for Medical Engineering and Science and Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[6]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, Massachusetts 02115, USA

[7]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

(Received 24 February 2025; published 15 July 2025)

Antimicrobial resistance is a critical global health challenge and one of the World Health Organization's top ten public health threats. The alarming rise of drug-resistant pathogens threatens to usher in a postantibiotic era where common infections could once again become fatal. Despite the urgency, traditional discovery methods are time-consuming, expensive, and insufficient to keep pace with rapidly evolving resistance. Recent advances in machine learning (ML) and artificial intelligence (AI) present a transformative alternative, enabling the rapid identification of potential candidate antibiotics in a fraction of the time required by conventional methods. In this Essay, we discuss how ML and AI significantly accelerate antibiotic discovery, drawing on previous works and insights in this emerging field. We also consider the promises and challenges of this emerging area and speculate on its evolution in the coming years, highlighting the potential contributions from the physics community.

Part of a series of Essays in Physical Review Letters which concisely present author visions for the future of their field.

DOI: 10.1103/y3fg-s9vg

Introduction—Antimicrobial resistance (AMR) is an escalating global health crisis, with the World Health Organization labeling it a top health threat [1]. The rise of drug-resistant pathogens threatens to reverse decades of medical progress, transforming once-treatable infections into potentially lethal conditions [2,3]. Traditional antibiotic development methods often require years to produce preclinical candidates, making them increasingly inadequate to combat the rapid emergence of resistant strains. Currently, drug-resistant bacterial infections claim over 1.27 million lives annually, a figure projected to climb to 10 million deaths per year by 2050. This scenario could herald a postantibiotic era with devastating consequences. Global efforts have emphasized antibiotic stewardship, new therapeutic development, and enhanced diagnostics to combat this looming crisis. Yet the antibiotic discovery pipeline remains perilously thin, impeded by high development costs, lengthy timelines, and low returns on investment. Pharmaceutical companies have increasingly turned away from antibiotic R&D, further exacerbating the crisis.

Recent advances in machine learning (ML) and artificial intelligence (AI) are beginning to reshape the antibiotic discovery landscape [4–15]. Just a decade ago, many of these breakthroughs would have seemed impossible. Today, AI can rapidly parse massive datasets to find antibiotic molecules in both extant [11] and extinct organisms [9,10,16]. It can also predict the antibiotic properties of known compounds and repurpose them from vast compound libraries [8], a task that would be infeasible for humans to do manually at scale. By conceptualizing biology as information, AI has also allowed the rapid exploration and identification of new molecules with therapeutic potential [9–13]. Moreover, we are beginning to incorporate explainability into these models [14,17], enhancing our understanding of how they work, an aspect particularly relevant for deep learning models. This growing field has experienced numerous advances in recent years, auguring a promising future for the field of antimicrobial resistance [18–25]. Importantly, these

*Contact author: cfuente@upenn.edu
†Contact author: jimjc@mit.edu

computational advances have been validated experimentally, with AI-discovered molecules undergoing extensive *in vitro* and *in vivo* testing. Over the past six years, multiple machine-identified compounds have demonstrated safety and efficacy in preclinical mouse models, marking them as promising candidates for further development. Beyond its scientific promise, this Essay also reduces the environmental footprint of antibiotic R&D by shifting most screening in silico, curtailing the plastics, reagents, and energy required for wet-lab iterations.

In this Essay, we highlight the emerging field of AI-driven antibiotic discovery and how this convergence of computational power, biological data, and interdisciplinary expertise from biology, physics, and chemistry holds great potential to address AMR.

Biology as information and AI for antibiotic discovery—Information theory, pioneered by Shannon [26], transformed our understanding of how information can be encoded, transmitted, and processed. Similarly, biology can be conceptualized as a massive information repository, stored in genomes and proteomes (i.e., the full set of proteins encoded in a genome). Within this framework, it is conceivable to develop algorithms capable of mining this biological information to uncover new, useful molecules such as antibiotics, following workflows illustrated in Fig. 1. These AI-driven approaches use computational models to identify potential antibiotic molecules, which are then synthesized and validated through cytotoxicity tests, antimicrobial assays, mechanism of action studies, and safety and efficacy evaluations in animal models. Historically, antibiotic discovery depended on laborious, trial-and-error strategies, such as bioprospecting—collecting samples from nature, isolating active compounds, and testing them for efficacy. This process is slow, costly, and unpredictable, often yielding few viable candidates.

Recognizing these limitations, we hypothesized years ago that machine learning methods could offer a more systematic and expedited approach to antibiotic design and discovery [4], and that would dramatically shorten the timeline for identifying new compounds compared to traditional methods. We proposed leveraging decades of accumulated biological data—such as genomes and proteomes—that, with algorithms devised to identify active compounds, could be systematically searched to discover novel antibiotics [11]. Early proof-of-concept work focused on identifying antibiotic sequences in individual proteins [28]. Proteins, like all biomolecules, can be viewed as biological codes composed of amino acid strings, and the hypothesis was that among these amino acid sequences lay hidden segments with antimicrobial properties. The advent of improved computational resources and sophisticated algorithms has expanded these initial efforts to entire proteomes. For instance, large-scale exploration of the human proteome uncovered thousands of previously unrecognized antimicrobial sequences, some likely involved in host immunity [11].

Encouraged by these results, we then reasoned that such antimicrobial sequences might be ubiquitous across the tree of life, including in bacterial genomes. Indeed, recent efforts uncovered nearly one million antibiotic candidates by analyzing microbial "dark matter" [13] and human microbiomes [29]. Gene prediction tools, such as Prodigal [30] and Macrel [31], which accurately predict and classify small open reading frames (smORFs, see Fig. 1), enabled the systematic mining of 2.7 billion smORFs, yielding a vast catalog of potential antibiotics. Additional computational approaches have similarly identified antimicrobial agents in diverse microbiomes [32].

These types of computational methods have dramatically accelerated the discovery of new antibiotics. Once a years-long process, antibiotic discovery can now be accomplished in a matter of hours [6,33], highlighting how AI has fundamentally transformed the field. As these methods continue to improve, it is reasonable to expect that a substantial portion of the world's biological data will soon be mined for antibiotic molecules, ultimately aiming to move some of these discoveries into human clinical trials. However, the success of AI models in antibiotic discovery hinges on the quality, size, and diversity of the training data [34]. Generating comprehensive datasets is expensive and time intensive, often spanning multiple years. Existing biological datasets, including those pertinent to infectious diseases and antibiotic discovery, are frequently proprietary, small, or poorly annotated. Overcoming these data barriers will require substantial investment and collaboration. To address these challenges, increased investment in dataset-generation projects is urgently needed. For example, APEX [10] was trained on a rigorously curated experimental dataset, including extensive negative controls, collected under standardized conditions over several years.

Collaborative efforts to create large, standardized, and publicly accessible datasets would greatly enhance the capability of AI models to predict novel antibiotics. Moreover, integrating diverse data types—such as genomic, proteomic, structural, and phenotypic data—could lead to more robust models capable of capturing the multifaceted nature of biological systems.

Looking forward, advancements in automation and high-throughput screening technologies [35] will play a significant role in streamlining the generation of high-quality data. The development of standardized protocols and data-sharing platforms, such as federated learning, will facilitate collaboration across disciplines and institutions, accelerating progress in the field. Furthermore, incorporating negative data—compounds tested but found inactive, which have been historically discarded—will be essential for yielding balanced datasets for training and benchmarking machine learning models.

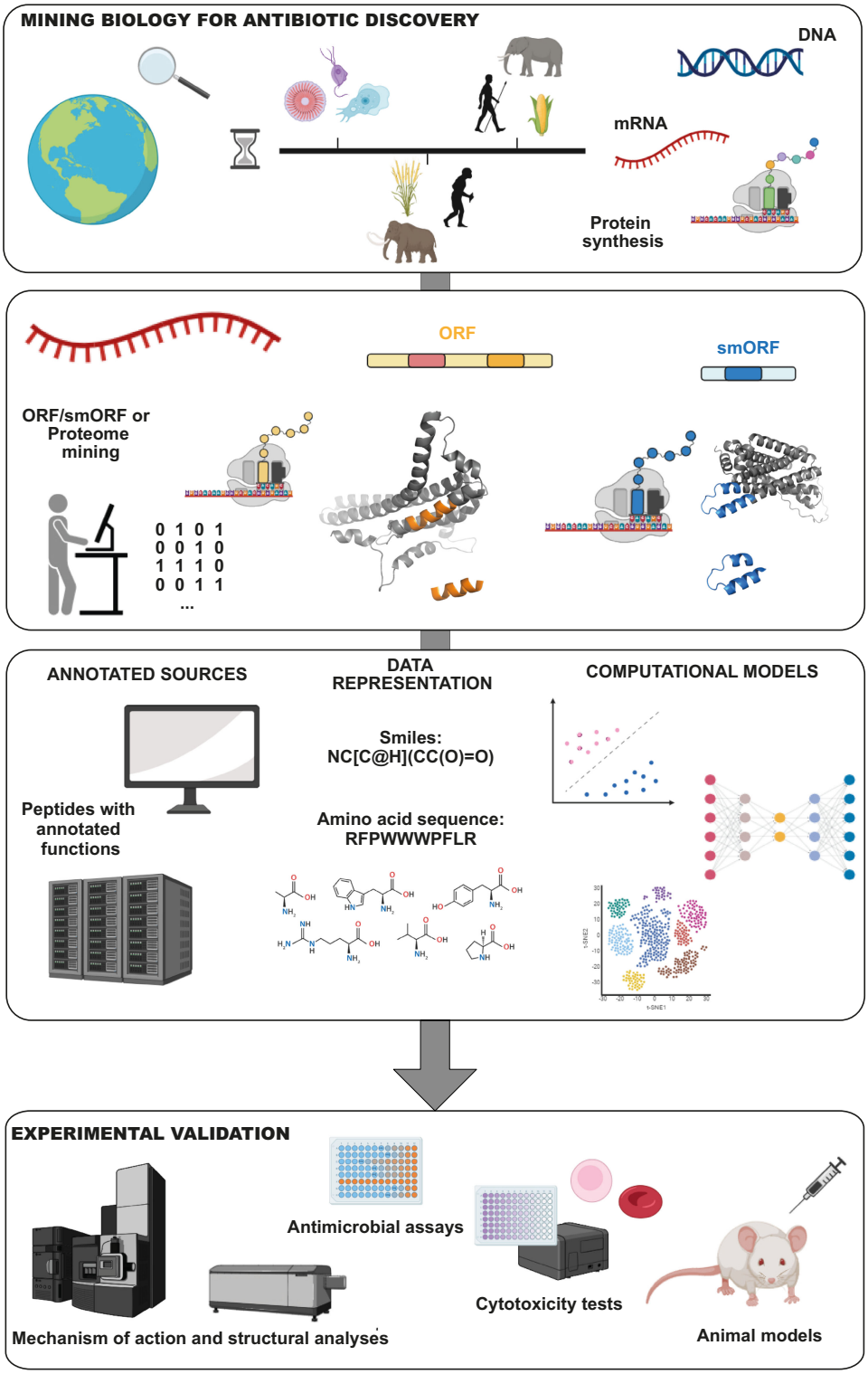


FIG. 1. Example of AI-driven antibiotic discovery. This figure illustrates the workflow involved in mining biological data (e.g., from genomes, proteomes, and metagenomes) to identify and validate antibiotics with annotated functions. For example, computational and AI models can be used to mine the proteome, leading to the annotation of peptides and their sequences, such as the example of a simplified molecular-input line-entry system (SMILES) code and an amino acid sequence shown in the figure. SMILES is a text-based format used to represent chemical structures, encoding atoms, bonds, and stereochemistry with specific symbols to facilitate easy storage and computational analysis. The annotated molecules are then synthesized and undergo experimental validation through various assays, including cytotoxicity tests, antimicrobial assays, mechanism of action studies, structural analyses, and safety and efficacy studies using animal models. Figure created with [27].

Molecular deextinction: A new framework for antibiotic discovery—Having discovered numerous antibiotic sequences in the human proteome, we ventured further into evolutionary history, hypothesizing that our extinct relatives, such as Neanderthals and Denisovans, might harbor similarly bioactive molecules. By analyzing their proteomes, we uncovered novel antimicrobial sequences and introduced the concept of *molecular deextinction* [9,10]. This approach aims to identify ancient molecular sequences to understand their evolution and tackle modern challenges like AMR, positing that molecules from extinct species may be effective against contemporary pathogens. To scale this idea beyond a few extinct proteomes to a large *extinctome* of all known extinct organisms, we developed a more advanced deep-learning model called APEX [10]. This model combines recurrent neural networks (RNNs) with attention mechanisms and is trained on experimentally validated data. APEX has successfully identified antibiotic molecules derived from various ancient creatures, including the woolly mammoth. Some of these molecules, such as neanderthalin, mammuthusin, and elephasin, have already shown promise in preclinical testing.

Over the next decade, molecular deextinction could redefine our understanding of peptide and protein evolution, providing insights into how mutations over time influenced biological function. These ancient sequences could inform new therapeutic strategies and advance research in bioinformatics, physiology, immunology, and more [36]. These approaches would benefit from advanced sequence-to-structure prediction tools (like AlphaFold) and high-throughput validation pipelines to confirm the reconstructed proteins' functions. In principle, the same framework could be adapted to study viral evolution: By analyzing historical or ancestral viral sequences, ML models could help predict the structural and functional changes that may give rise to novel pathogens, much like COVID-19, and guide preemptive research and drug design efforts.

Interpretability—A major challenge in AI-driven drug discovery is the black-box nature of many algorithms, particularly deep learning models. Developing explainable AI models is crucial for advancing the field and will allow for the refinement of models based on a clearer understanding of their outputs. In recent work [14], we developed an explainable, substructure-based deep learning approach to explore chemical spaces, predicting antibiotic activity and cytotoxicity for over 12 million compounds.

In this context, explainable refers to the potential to analyze and understand the predictions of graph neural networks in the context of drug discovery by identifying chemical substructures, or *rationales*, that influence the model's predictions. This allows understanding of which specific molecular features or patterns contribute to a compound's predicted antibiotic activity, making the model's outputs more transparent and interpretable. This method led to the identification of novel structural classes, including one that is effective against methicillin-resistant *Staphylococcus aureus* and vancomycin-resistant enterococci, by revealing structural motifs linked to antibiotic activity. These chemical insights, such as disrupting the proton motive force by altering pH gradients, demonstrate how machine learning can uncover targeted mechanisms for antibiotic discovery.

Explainable ML approaches are critical for capturing the specific aspects of training data that models have learned, either by identifying input structural features or by highlighting model components [37], leading to a particular prediction. Active research is increasingly focused on hybrid approaches that combine experimental and computational methods to enhance predictive accuracy and address data scarcity. ML techniques that effectively utilize limited training data while exploring extensive chemical search spaces are likely to make significant contributions to anti-infective drug discovery. Particularly for complex models like graph neural networks, explainable ML approaches offer useful tools for identifying critical structural features that drive model predictions. Computational pipelines that incorporate structural predictions of proteins and other macromolecules can further enhance model accuracy. Techniques such as molecular dynamics simulations [38] and ML-augmented docking strategies [39] exemplify methods that can improve predictions of drug-macromolecule interactions. Additionally, sequence-to-structure models, like AlphaFold for proteins or Rhofold+ for RNAs, will likely play an important role in structure-guided drug design. These models can be used to fine-tune therapeutic candidates to achieve specific structural conformations, thereby bridging structural predictions with the productive expansion of chemical search spaces.

In the future, developing more interpretable models will be essential for the regulatory acceptance and clinical adoption of AI-discovered drugs. Techniques such as attention mechanisms, feature attribution methods, and model distillation can significantly enhance model interpretability. Additionally, integrating domain knowledge into AI models will improve their transparency and reliability, fostering greater trust among regulators and clinicians.

Complementary physics-based methods for better predictability: While ML has opened new horizons in antibiotic discovery, physics-based methods offer additional insights that can significantly enhance the predictive power and practical impact of computational models. Techniques such as molecular dynamics (MD) simulations and quantum mechanics or molecular mechanics methods provide detailed atomic-level insights into drug-target molecular interactions [40,41]. Incorporating these simulations with ML models can help improve our understanding of binding affinities, reaction mechanisms, and

enzyme activities critical to designing novel antimicrobials [6].

Traditional ML models often rely on static representations of biological systems. In contrast, MD simulations allow us to transition from a static understanding of these systems to modeling the dynamic behavior of biomolecules over time, capturing conformation changes and interactions that are not apparent from static structures [40,42]. Integrating MD simulations with ML algorithms may help improve the prediction of binding affinities and identify potential off-target effects.

Quantum mechanical methods, such as density functional theory (DFT) may also be integrated, as they enable precise calculations of electronic structures and can be used to model chemical reactions at the active sites of enzymes [41]. These hybrid simulations could help design inhibitors targeting specific enzymatic pathways central to antibiotic resistance, such as β-lactamases or ribosomal modification enzymes.

In terms of experimental validation, advances in automation and microfluidics, both of which are directly influenced by physics, can facilitate high-throughput screening of AI-predicted candidates [43]. Tiny assay volumes reduce reagent costs and enable dynamic, real-time monitoring of bacterial growth or viability. Organ-on-a-chip systems, which mimic physiological conditions, can yield more clinically relevant data compared to standard *in vitro* assays, thereby improving model accuracy. Closed-loop microfluidic setups, where ML predictions guide experiments and experimental data refine ML models, represent a promising path forward.

Collectively, these physics-based tools bridge the gap between computational predictions and real-world validation, ensuring that ML-driven antibiotic discovery extends beyond algorithmic promise to tangible clinical outcomes. Future developments in computational power and algorithms are expected to make these physics-based simulations more accessible and faster, allowing them to be integrated more seamlessly with ML approaches. The continued collaboration between physicists, chemists, biologists, and computer scientists will be essential to fully realize the potential of these interdisciplinary strategies in combating AMR.

Future challenges and opportunities—For many years, the idea that AI and computers could be successfully applied to biology, including in fields such as microbiology and antibiotic discovery, was met with skepticism. While AI found success in fields like image and speech recognition, the prevailing notion was that the complexity of biology was beyond the reach of machines. However, recent advancements have demonstrated the potential of ML to dramatically accelerate antibiotic discovery, opening new avenues to addressing one of humanity's greatest challenges, antimicrobial resistance. This field is inherently transdisciplinary, drawing scientists from a wide range of disciplines. This convergence of expertise has been a key factor in the successes achieved so far and will continue to be critical as AI technologies for antibiotic discovery evolve and improve.

Looking ahead, we anticipate significant progress in several key areas. The availability of large, standardized datasets will provide the foundational training sets necessary for improving AI models, with automation likely playing a major role in streamlining the generation of these high-quality datasets. Over the next decade, we expect the emergence of multimodal algorithms capable of simultaneously optimizing multiple parameters, such as antibiotic activity, while minimizing off-target effects like cytotoxicity, hemolysis, and disruption of beneficial microbiota. These algorithms will also optimize for stability, pharmacokinetics or pharmacodynamics profiles, and other critical drug development parameters. They will be instrumental in designing compounds with favorable druglike properties, accelerating the pipeline from discovery to clinical application.

With continued advancements in computational power, we envision the possibility of fully exploring all of the world's biological information as a digital resource for discovering new antibiotics. Innovations in generative models, such as large language models, variational autoencoders (VAEs), and generative adversarial networks, will further accelerate our ability to explore vast chemical spaces and create novel biomolecular structures with desired properties. AI will also play a crucial role in refining existing antibiotics by optimizing their structures to enhance efficacy and reduce toxicity. A practical way to refine existing antibiotics would be to combine AI-driven candidate generation with quantum chemistry methods like DFT to evaluate electronic structures and reaction pathways. First, AI models (e.g., deep learning or generative algorithms) would suggest chemical modifications aimed at enhancing antibiotic potency and reducing toxicity. Then, quantum chemical calculations could verify the most promising leads by accurately predicting binding energies, conformational changes, and potential off-target effects at the atomic level.

Moreover, algorithms will be instrumental in decoding and predicting resistance mechanisms. By analyzing genomic and phenotypic data, computational tools can be used to identify potential resistance pathways and predict the emergence of resistance [44]. These models can also scrutinize genomic data from bacterial populations to detect antibiotic-resistance markers, guiding the design of drugs that can circumvent known resistance mechanisms.

Finally, AI may also be employed to identify synergistic combinations of antibiotics that enhance efficacy and reduce the likelihood of resistance. For example, by scrutinizing drug interaction, these models can predict which compound cocktails are the most effective against resistant strains, potentially yielding novel combination therapies that outperform single-drug treatments.

In conclusion, AI holds immense promise in the fight against AMR. While there are still limitations and challenges to overcome, we are optimistic about the progress made so far and the potential for AI to deliver life-saving antibiotics in the future.

Acknowledgments—C. d. l. F.-N. is funded by a Presidential Professorship at the University of Pennsylvania, and the Langer Prize by the AIChE Foundation, and acknowledges funding from the IADR Innovation in Oral Care Award, the Procter & Gamble Company, United Therapeutics, a BBRF Young Investigator Grant, the Nemirovsky Prize, Penn Health-Tech Accelerator Award, the Dean's Innovation Fund from the Perelman School of Medicine at the University of Pennsylvania, the National Institute of General Medical Sciences of the National Institutes of Health under Grant No. R35GM138201, and the Defense Threat Reduction Agency (DTRA; HDTRA1-22-10031, HDTRA1-21-1-0014, HDTRA1-23-1-0001, HDTRA1-22-1-0032, and HDTRA1-22-1-0010). This work is also part of the Antibiotics-AI Project, which is directed by J. J. C. and supported by the Audacious Project, Flu Lab, LLC, the Sea Grape Foundation, R. Zander and H. Wyss for the Wyss Foundation, and an anonymous donor. We thank Angela Cesaro and Marcelo Torres for their assistance with the figure. We thank de la Fuente Lab members for insightful discussions.

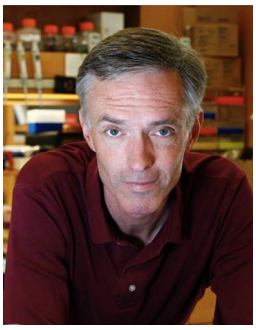
Conceptualization, writing, and editing: C. d. l. F.-N. and J. J. C.

C. d. l. F.-N. is a co-founder of, and scientific advisor, to Peptaris, Inc., provides consulting services to Invaio Sciences, and is a member of the Scientific Advisory Boards of Nowture S.L., Peptidus, European Biotech Venture Builder, the Peptide Drug Hunting Consortium (PDHC), ePhective Therapeutics, Inc., and Phare Bio. provides consulting services to Invaio Sciences and is a member of the Scientific Advisory Boards of Nowture S. L., Peptidus, and Phare Bio. He is also on the Advisory Board of the Peptide Drug Hunting Consortium (PDHC). J. J. C. is a co-founder and board member of Phare Bio, a nonprofit focused on antibiotic drug development.

Data availability—No data were created or analyzed in this Essay.

César de la Fuente

Cesar de la Fuente-Nunez is a Presidential Associate Professor at the University of Pennsylvania. He is best known for pioneering computational and artificial intelligence approaches to antibiotic discovery, which have drastically accelerated the time needed to identify preclinical candidates, showing promise for therapeutic intervention against currently untreatable infections. Professor de la Fuente-Nunez has received numerous awards and honors for his contributions, including the Princess of Girona Prize, the Miklós Bodanszky Award, and the Fleming Prize. He is a Fellow of the American Institute for Medical and Biological Engineering, a Sloan Fellow, and a National Academy of Medicine Emerging Leader. He has authored more than 170 publications and holds multiple patents.



James J. Collins

James J. Collins is the Termeer Professor of Medical Engineering and Science and Professor of Biological Engineering at MIT, as well as a Member of the Harvard-MIT Health Sciences and Technology Faculty. He is a Core Founding Faculty member of the Wyss Institute for Biologically Inspired Engineering at Harvard University and an Institute Member of the Broad Institute of MIT and Harvard. He is one of the founders of the field of synthetic biology, and his research group is currently focused on using synthetic biology to create next-generation diagnostics and therapeutics. J. J. C. is the Director of the Antibiotics-AI Project at MIT and co-founder of Phare Bio, a nonprofit focused on AI-driven antibiotic discovery. He has received numerous awards and honors, including a MacArthur "Genius" Award, the Max Delbruck Prize in Biological Physics, and the Feynman Prize in Nanotechnology, and he is an elected member of all three national academies: the National Academy of Sciences, the National Academy of Engineering, and the National Academy of Medicine.

[1] EClinicalMedicine, Antimicrobial resistance: A top ten global public health threat, EClinicalMedicine **41**, 101221 (2021).

[2] C. de la Fuente-Nunez, A. Cesaro, and R. E. W. Hancock, Antibiotic failure: Beyond antimicrobial resistance, Drug Resist. Updates **71**, 101012 (2023).

[3] K. Lewis, The science of antibiotic discovery, Cell **181**, 29 (2020).

[4] M. T. Torres and C. de la Fuente-Nunez, Toward computer-made artificial antibiotics, Curr. Opin. Microbiol. **51**, 30 (2019).

[5] M. C. R. Melo, J. R. M. A. Maasch, and C. de la Fuente-Nunez, Accelerating antibiotic discovery through artificial intelligence, Commun. Biol. **4**, 1050 (2021).

[6] F. Wong, C. de la Fuente-Nunez, and J. J. Collins, Leveraging artificial intelligence in the fight against infectious diseases, Science **381**, 164 (2023).

[7] W. F. Porto, L. Irazazabal, E. S. F. Alves, S. M. Ribeiro, C. O. Matos, Á. S. Pires, I. C. M. Fensterseifer, V. J. Miranda, E. F. Haney, V. Humblot, M. D. T. Torres, R. E. W. Hancock, L. M. Liao, A. Ladram, T. K. Lu, C. de la Fuente-Nunez, and O. L. Franco, In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design, Nat. Commun. **9**, 1490 (2018).

[8] J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carfrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, and J. J. Collins, A deep learning approach to antibiotic discovery, Cell **180**, 688 (2020); **181**, 475(E) (2020).

[9] J. R. M. A. Maasch, M. D. T. Torres, M. C. R. Melo, and C. de la Fuente-Nunez, Molecular de-extinction of ancient antimicrobial peptides enabled by machine learning, Cell Host Microbe **31**, 1260 (2023).

[10] F. Wan, M. D. T. Torres, J. Peng, and C. de la Fuente-Nunez, Deep-learning-enabled antibiotic discovery through molecular de-extinction, Nat. Biomed. Eng. **8**, 854 (2024).

[11] M. D. T. Torres, M. C. R. Melo, L. Flowers, O. Crescenzi, E. Notomista, and C. de la Fuente-Nunez, Mining for encrypted peptide antibiotics in the human proteome. Nat. Biomed. Eng. **6**, 67 (2022).

[12] A. Cesaro, M. D. T. Torres, R. Gaglione, E. Dell'Olmo, R. Di Girolamo, A. Bosso, E. Pizzo, H. P. Haagsman, E. J. A. Veldhuizen, C. de la Fuente-Nunez, and A. Arciello, Synthetic antibiotic derived from sequences encrypted in a protein from human plasma, ACS Nano **16**, 1880 (2022).

[13] C. D. Santos-Júnior, M. D. T. Torres, Y. Duan, Á. Rodríguez Del Río, T. S. B. Schmidt, H. Chong, A. Fullam, M. Kuhn, C. Zhu, A. Houseman, J. Somborski, A. Vines, X. M. Zhao, P. Bork, J. Huerta-Cepas,C. de la Fuente-Nunez, and L. P. Coelho, Discovery of antimicrobial peptides in the global microbiome with machine learning, Cell **187**, 3761 (2024).

[14] F. Wong *et al.*, Discovery of a structural class of antibiotics with explainable deep learning, Nature (London) **626**, 177 (2024).

[15] G. Liu, D. B Catacutan, K. Rathod, K. Swanson, W. Jin, J. C. Mohammed, A. Chiappino-Pepe, S. A. Syed, M. Fragis, K. Rachwalski, J. Magolan, M. G. Surette, B. K. Coombes, T. Jaakkola, R. Barzilay, J. J. Collins, and J. M. Stokes, Deep learning-guided discovery of an antibiotic targeting *Acinetobacter baumannii*, Nat. Chem. Biol. **19**, 1342 (2023).

[16] F. Wan, M. D. T. Torres, J. Peng, and C. de la Fuente-Nunez, Molecular de-extinction of antibiotics enabled by deep learning, bioRxiv (2023), 10.1101/2023.10.01.560353.

[17] B. Wang, P. Lin, Y. Zhong, X. Tan, Y. Shen, Y. Huang, K. Jin, Y. Zhang, Y. Zhan, D. Shen, M. Wang, Z. Yu, and Y. Wu, Explainable deep learning and virtual evolution identifies antimicrobial peptides with activity against multidrug-resistant human pathogens, Nat. Rev. Microbiol. **10**, 332 (2025).

[18] P. Szymczak, M. Możejko, T. Grzegorzek, R. Jurczak, M. Bauer, D. Neubauer, K. Sikora, M. Michalski, J. Sroka, P. Setny, W. Kamysz, and E. Szczurek, Discovering highly potent antimicrobial peptides with deep generative model HydrAMP, Nat. Commun. **14**, 1453 (2023); **14**, 5129(E) (2023).

[19] G. R. Madden, R. H. Boone, E. Lee, C. D. Sifri, and W. A. Petri, Jr., Predicting *Clostridioides difficile* infection outcomes with explainable machine learning, EBioMedicine **106**, 105244 (2024).

[20] M. Njirjak, L. Žužić, M. Babić *et al.*, Reshaping the discovery of self-assembling peptides with generative AI guided by hybrid deep learning, Nat. Mach. Intell. **6**, 1487 (2024).

[21] Z. Chen, C. Ji, W. Xu, J. Gao, J. Huang, H. Xu, G. Qian, and J. Huang, UniAMP: enhancing AMP prediction using deep neural networks with inferred information of peptides, BMC Bioinf. **26**, 10 (2025).

[22] J. Yang, Y. Ran, S. Liu, C. Ren, Y. Lou, P. Ju, G. Li, X. Li, and D. Zhang, Synergistic D-amino acids based antimicrobial cocktails formulated via high through put screening and machine learning, Adv. Sci. **11**, e2307173 (2024).

[23] H. Zhang, Y. Wang, Y. Zhu, P. Huang, Q. Gao, X. Li, Z. Chen, Y. Liu, J. Jiang, Y. Gao, J. Huang, and Z. Qin, Machine learning and genetic algorithm-guided directed evolution for the development of antimicrobial peptides, J. Adv. Res. **68**, 415 (2025).

[24] L. Yao, J. Guan, P. Xie, C. R. Chung, J. Deng, Y. Huang, Y. C. Chiang, and T. Y. Lee, AMPActiPred: A three-stage framework for predicting antibacterial peptides and activity levels with deep forest, Protein Sci. **33**, e5006 (2024).

[25] Y. Ma, Z. Guo, B. Xia *et al.*, Identification of antimicrobial peptides from the human gut microbiome using deep learning, Nat. Biotechnol. **40**, 921 (2022).

[26] C. E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. **27**, 379 (1948).

[27] https://www.biorender.com/

[28] K. Pane, V. Cafaro, A. Avitabile, M. T. Torres, A. Vollaro, E. De Gregorio, M. R. Catania, A. Di Maro, A. Bosso, G. Gallo, A. Zanfardino, M. Varcamonti, E. Pizzo, A. Di Donato, T. K. Lu, C. de la Fuente-Nunez, and E. Notomista, Identification of novel cryptic multifunctional antimicrobial peptides from the human stomach enabled by a computational-experimental platform, ACS Synth. Biol. **7**, 2105 (2018).

[29] M. D. T. Torres, E. F. Brooks, A. Cesaro, H. Sberro, M. O. Gill, C. Nicolaou, A. S. Bhatt, and C. de la Fuente-Nunez, Mining human microbiomes reveals an untapped source of peptide antibiotics, Cell **187**, 5453 (2024).
[30] D. Hyatt, G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser, Prodigal: Prokaryotic gene recognition and translation initiation site identification, BMC Bioinf. **11**, 119 (2010).
[31] C. D. Santos-Júnior, S. Pan, X. M. Zhao, and L. P. Coelho, Macrel: Antimicrobial peptide screening in genomes and metagenomes, PeerJ Comput. Sci. **8**, e10555 (2020).
[32] Y. Ma, Z. Guo, B. Xia, Y. Zhang, X. Liu, Y. Yu, N. Tang, X. Tong, M. Wang, X. Ye, J. Feng, Y. Chen, and J. Wang, Identification of antimicrobial peptides from the human gut microbiome using deep learning, Nat. Biotechnol. **40**, 921 (2022).
[33] F. Wan, F. Wong, J. J. Collins *et al.*, Machine learning for antimicrobial peptide identification and design, Nat. Rev. Bioeng. **2**, 392 (2024).
[34] C. de la Fuente-Nunez, AI in infectious diseases: The role of datasets, Drug Resist. Updates **73**, 101067 (2024).
[35] M. Abolhasani and E. Kumacheva, The rise of self-driving labs in chemical and materials sciences, Nat. Synth. Gas **2**, 483 (2023).
[36] A. W. Torrance and C. de la Fuente-Nunez, The patentability and bioethics of molecular de-extinction, Nat. Biotechnol. **42**, 1179 (2024).
[37] J. H. Yang, S. N. Wright, M. Hamblin, D. McCloskey, M. A. Alcantar, L. Schrübbers, A. J. Lopatkin, S. Satish, A. Nili, B. O. Palsson, G. C. Walker, and J. J. Collins, A white-box machine learning approach for revealing antibiotic mechanisms of action, Cell **177**, 1649 (2019).
[38] P. Das, T. Sercu, K. Wadhawan, I. Padhi, S. Gehrmann, F. Cipcigan, V. Chenthamarakshan, H. Strobelt, C. Dos Santos, P. Y. Chen, Y. Y. Yang, J. P. K. Tan, J. Hedrick, J. Crain, and A. Mojsilovic, Accelerated antimicrobial discovery via deep generative models and molecular dynamics simulations, Nat. Biomed. Eng. **5**, 613 (2021); **5**, 942 (2021).
[39] F. Wong, A. Krishnan, E. J. Zheng, H. Stärk, A. L. Manson, A. M. Earl, T. Jaakkola, and J. J. Collins, Benchmarking AlphaFold-enabled molecular docking predictions for antibiotic discovery, Mol. Syst. Biol. **18**, e11081 (2022).
[40] S. A. Hollingsworth and R. O. Dror, Molecular dynamics simulation for all, Neuron **99**, 1129 (2018).
[41] H. M. Senn and W. Thiel, QM/MM methods for biomolecular systems, Angew. Chem., Int. Ed. Engl. **48**, 1198 (2009).
[42] A. Talevi, Computer-aided drug discovery and design: Recent advances and future prospects, Methods Mol. Biol. **2714**, 1 (2024).
[43] E. K. Sackmann, A. L. Fulton, and D. J. Beebe, The present and future role of microfluidics in biomedical research, Nature (London) **507**, 181 (2014).
[44] Ö. Baltekin, A. Boucharin, E. Tano, D. I. Andersson, and J. Elf, Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging, Proc. Natl. Acad. Sci. U.S.A. **114**, 9170 (2017).