



Perspective

An aurora of natural products-based drug discovery is coming

Yaojun Tong^{c,*}, Zixin Deng^{a,b,**}

^a State Key Laboratory of Microbial Metabolism, School of Life Science and Biotechnology, Shanghai Jiao Tong University, Shanghai, 200030, China

^b Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and School of Pharmaceutical Sciences, Wuhan University, Wuhan, 430071, China

^c The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800, Kgs. Lyngby, Denmark

A B S T R A C T

Natural products (NPs), a nature's reservoir possessing enormous structural and functional diversity far beyond the current ability of chemical synthesis, are now proving themselves as most wonderful gifts from mother nature for human beings. Many of them have been used successfully as medicines, as well as the most important sources of drug leads, food additives, and many industry relevant products for millennia. Most notably, more than half of the antibiotics and anti-cancer drugs currently in use are, or derived from, natural products. However, the speed and outputs of NP-based drug discovery has been slowing down dramatically after the fruitful harvest of the

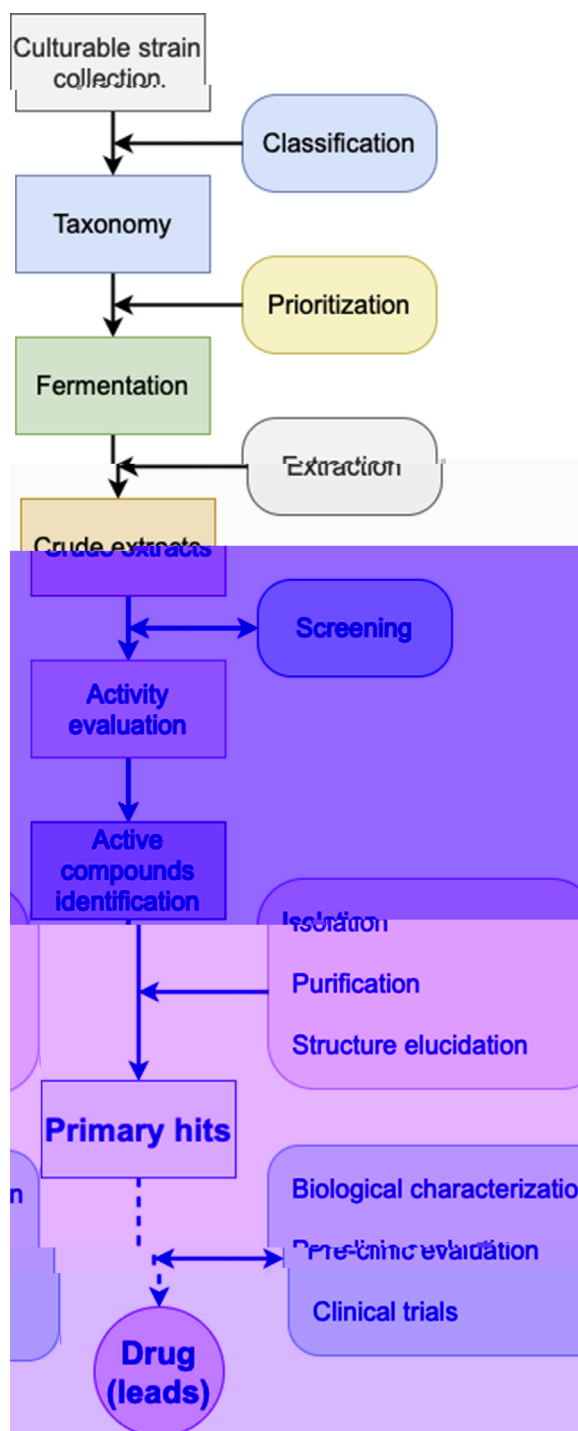


Fig. 1. A simple workflow of the “top-down” strategy for natural products discovery. Dash line means uncertainty, indicating there is a risk of failure from hits to drugs. Double-headed arrow means a feedback loop may exist.

30 years now with no new approvals of natural product-based antibiotics [9]. With the emergence of combinatorial synthesis, a large number of synthetic libraries have been synthesized. It made many pharmaceutical companies step back and even exit from the natural products arena, which resulted in a sharp reduction in the output of both new drug leads and approved drugs from the drug development pipeline [10]. However, with the advances in whole genome sequencing and genome mining, we see that there is a huge yet unexploited potential of secondary metabolites which is hidden in the genome of the microorganisms, both culturable and unculturable [11]. Typically,

there are 20–50 biosynthetic gene clusters (BGCs) encoding secondary metabolites per actinomycete strain, less than 10% of which have been chemically identified in laboratory conditions. Given the fact that new bioactivities are inherently linked with novel chemical structures, we could imagine a large number of new bioactive secondary metabolites are waiting on us to discover them. The award of the 2015 Nobel Prizes in Physiology or Medicine for research in natural products both validated and reinvigorated the whole natural product community [12]. The incredible rate of development in genome sequencing, modern metabolic engineering, synthetic biology, advanced genome editing, big data, artificial intelligence (AI), and 3D printing together with the growing microbial strain collections enable us to access the previously inaccessible natural products. Taking these advances all together, there is no doubt that a new age of natural products discovery is yet to come.

Personal perspectives

Resources and natural products

As the Chinese proverb goes, “without rice, even the cleverest housewife cannot cook”. Natural products discovery also requires good resources. The resources discussed here are mainly microbial strain collections and high-quality genome/metagenome sequences. Microbes, as the natural products producers, are the start point of the whole journey of microbial natural products discovery, making themselves irreplaceable. Initiated by Professor V. B. D. Skerman, the World Federation for Culture Collection (WFCC)-World Data Centre for Microorganisms (WDCM) has been growing into WFCC-MIRCEN (Microbial Resources Centres)-WDCM, the world’s leading data center for microbial resources (<http://www.wdcm.org>). It has been hosted and maintained by the Institute of Microbiology, Chinese Academy of Sciences since 2010. To date, more than 3,233,000 microbials from 791 culture collections in 78 countries were registered in WFCC-MIRCEN-WDCM. Beyond these registered resources, there also are some well-established microbial strain collections in both non-profit research institutes and companies (Table 1).

Given the fact that the majority of microorganisms are unculturable with current technologies, genomic and metagenomic sequence information are becoming more and more important as nowadays we are able to directly get encoded compounds from the genetic information. A detailed introduction of microbial genome resources can be found in the review [13]. Two such well-known databases are GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) and Joint Genome Institute (JGI) Genome Portal (<http://genome.jgi.doe.gov/portal/>). To date, they have recorded > 21,000 and > 16,000 complete bacterial genomes, respectively. It’s worth noting that a large part of these complete genomes are resequencing results of the same species. Only around 260 complete genomes for streptomycetes are available in GenBank. Of course, many genome sequencing projects are ongoing, for example the 10 k microbial sequencing project coordinated by Dr. Lixin Zhang, and also many microbial genome sequences have not yet been deposited to public databases, for example the genomic database of over 135,000 strains that was originally assembled by Warp Drive Bio, and now it belongs to Ginkgo Bioworks.

Synthetic biology and natural products

With rice ready, now we need tools to cook the meals. Tools related to natural products discovery will be discussed in following sections.

After the discovery the double helix of DNA and the human genome project (HGP), synthetic biology was crowned “the third biotechnology revolution”. It is a product of multi-disciplinary integration, which is one of the most active areas in biological and biotechnological development. This highly interdisciplinary area applies engineering principles to biology. It involves biology, evolution, chemistry, physics, mathematics, engineering, and informatics. Applications of synthetic

biology have made significant achievements in many areas, such as bioenergy, biomaterials, biomedicine, and bulk chemicals. It greatly enables us to better understand and even engineer life. Given the complexity of the biosynthetic pathway of natural products, the “top-down” strategy is obsolete, the more acceptable way to study natural products is the “bottom-up”

product are highly linked to the whole metabolic network (primary and secondary metabolisms) [16]. We have to know which pathways/enzymes have crosstalk; when, where, and how many enzymes are needed; how to balance the cell growth and production; how to direct the metabolic flow to the target pathway to reach the theoretical yield. A classic example is the production of penicillin, the yield increased more than 10,000 times by simple strain (the cell factory) improvement. With the continuous development in CRISPR-based genome editing techniques, such as CRISPR-Cas9 [17], CRISPR base editor [18,19], and CRISPR prime editing [20] enables much faster and more detailed strain improvement.

Given the fact that most of the microbial gene clusters encoding secondary metabolites are so called “cryptic/silent” gene clusters, named for their non-expression and/or trace expression, heterologous expression becomes a powerful approach. With the growth of the metabolism knowledge base, and with the advances in genome editing and metabolic modeling, we are now able to design and construct better microbial cell factories for activation and/or high production of natural products. For example, the anti-malaria drug artemisinin was primarily from *A. annua*. Due to the low yield, seasonal and regional limitations, it could not meet the market demand. However, artemisinin can easily be chemically synthesized from the precursor artemisinic acid, therefore, researchers reconstituted the biosynthetic pathway of artemisinic acid in a yeast cell factory, the yield can reach as high as 25 g/L in fermentation [21]. Similarly, the bonds between opioid production and requiring field-grown poppies were broken. Recently, researchers successfully produced hydrocodone in a yeast cell factory by reconstitution of a complete hydrocodone biosynthetic pathway involving 23 enzymatic steps [22]. Besides yeast cell factories, like *Saccharomyces cerevisiae* and *Yarrowia lipolytica*, some other microbes have great potential for being cell factories in regards to natural product production, such as: *E. coli*, which has the most well established knowledge base and toolkit; *Pseudomonas putida*, which in general has high tolerance to many chemicals; and actinomycetes (*S. albus*, *S. coelicolor*, *S. avermitilis*, *S. ambofaciens*, *Saccharopolyspora erythraea*, etc), which could be more suitable for expressing biosynthetic gene clusters with high GC content. The design and construction of actinomycete cell factories have been heavily limited by the humble traditional genetic manipulation approaches, the recently established CRISPR based genome editing methods for actinomycetes [23–25] bring us possibilities to efficiently make good actinomycete cell factories. Moreover, the emerging of CRISPR-based biosensor development [26] enables faster and more sensitive detection of target compounds produced by microbial cell factories.

Reading and writing DNA and natural products

During the past 50 years, we have witnessed the incredible development of DNA sequencing technologies. The first whole genome to be sequenced was that of the bacteriophage ϕ X174 in 1977 [27], with only 5,386 bp. To get a human genome (3,234.83 Mb) sequenced, the HGP took 13 years (1990–2003), involved 20 universities in 6 countries, and spent ~3 billion US dollars. By comparison, the Broad Institute announced that it had sequenced 100,000 whole human genomes by 2018. With the fast evolution of sequencing related technologies, from shotgun sequencing, to pyrosequencing (454), to Illumina sequencing, and now to long read PacBio sequencing and Nanopore sequencing, DNA sequencing has become much cheaper, easier and orders of magnitude faster. We are now entering the age of the 1K genome (it only costs 1,000 US dollars for whole genome sequencing of a human).

The increased ability of DNA reading (sequencing) has made the information of DNA sequences indispensable for biological research. The exponentially accumulated information of whole genome sequences is also changing how natural products are discovered, instead of the traditional “top-down” approach, we are shifting to the genome mining based “bottom-up” strategy. Genome mining confirms that there

is still a huge potential of novel natural products in the microbial genomes [28].

Compared to the ability to read DNA, the ability to write DNA still has a long way to go. The current nucleic acid synthesis relies heavily on chemical synthesis, with relatively high cost and a long processing time. Can we develop a bio-based (mimicking living organisms), high-fidelity, cheap, and fast nucleic acid synthesis platform to reach the stage of “made-to-order” DNA for desired purposes? One bottle neck of the aforementioned “bottom-up” strategy is the cloning of long biosynthetic gene cluster with high GC content. One direct solution, of course, is the complete synthesis of the whole biosynthetic gene cluster and even the genome with reasonable cost and processing time [29].

Bioinformatics and natural products

We are accumulating tons and tons of data in every aspect. For example, the microbial whole genome sequences are becoming big data. For natural products discovery, the information we are looking for from the big data of genomes is clear, to find the right biosynthetic gene clusters with all necessary factors. However, these pieces of information are still like the needle in a haystack. Therefore, the processing of big data to determine the right way to find that needle is critical. More and more software and algorithms have been developed for natural products discovery. A good summary can be found in the Secondary Metabolite Bioinformatics Portal (<https://www.secondarymetabolites.org>) [30], currently it contains 25 such tools. The advances in AI and automation will also facilitate natural products discovery.

3D printing and natural products

3D printing is a process of building a three-dimensional object from a computer-aided design. It has been successfully applied in manufacturing, medical, industry, and many other areas. The list of these potential areas is growing. To date, 3D printing has not reached the molecular or even atomic level yet. Natural products-based drugs (or small molecules) that could be directly 3D printed on-demand would be highly advantageous. Some pioneer work has already been done recently [31]. If that day of 3D printing in the molecular level comes, and surely it will, it can help us to solve so many problems, such as the shortage of medical supplies in the fight against the sudden outbreak of some diseases such as COVID-19 caused by the coronavirus SARS-CoV-2.

Summary

We are in the best times, a lot of what was considered the impossible is becoming possible. Those previously inaccessible resources are becoming accessible. It is an age of pioneering and innovation. Worldwide scientific collaborations are extraordinarily frequent and easy. It is becoming more and more clear that multidisciplinary integration is the trend of scientific development, such as synthetic biology. We are the witnesses of history, but we also need to be prepared for being the builders of history. With world population growth increasing, areas such health, resources, and environment have drawn more and more attention. Natural products would be one of the keys to a bio-sustainable world. We rely on nature, learn from nature, and eventually we may outperform nature. Imagining a picture that in the near future, big data assists us in designing molecules for given diseases, AI helps us to construct the optimal biosynthetic pathways, which then will be synthesized by modern DNA synthesis, synthetic biology helps us to express the pathways and get the desired compounds. An even more wild thought would be to get the desired compounds directly by 3D printing.

Declaration of competing interest

There are no conflicts to declare.

Acknowledgements

The authors thank Dr. Wen-Jun Li and Dr. Lixin Zhang for agreeing to disclose the number of their microbial strain collections and providing unpublished information. The authors thank Simon Shaw for proofreading the manuscript. Y.T. acknowledges fundings from the Novo Nordisk Foundation (NNF10CC1016517; NNF15OC0016226; and NNF16OC0021746). Z.D. acknowledges funding from the National Natural Science Foundation of China (21661140002).

References

- [1] Borchardt JK. The beginnings of drug therapy: ancient mesopotamian medicine. *Drug News Perspect* 2002;15:187–92. <https://doi.org/10.1358/dnp.2002.15.3.840015>.
- [2] Huang KC. *The pharmacology of Chinese herbs*. second ed. ed. Boca Raton, Florida: CRC Press; 1998.
- [3] Fleming A. The discovery of penicillin. *Br Med Bull* 1944;2:4–5. <https://doi.org/10.1093/oxfordjournals.bmb.a071032>.
- [4] Waksman SA, Reilly HC, Johnstone DB. Isolation of streptomycin-producing strains of streptomyces-griseus. *J Bacteriol* 1946;52:393–7. <https://doi.org/10.1128/Jb.52.3.393-397.1946>