

# Discovery of New Compounds in Nature<sup>1</sup>

JON CLARDY

Professor of Biological Chemistry and Molecular Pharmacology  
Harvard Medical School

COMPOUNDS made by living creatures, especially the small molecules with pronounced biological activity, are some of the most remarkable objects on earth. They embody lessons about the diversity of molecular structures, about nature's strategy of creating small molecules, and about the ways that small molecules influence biological processes. These lessons have informed and inspired a large part of organic chemistry. They have also provided biologists with tools to probe and understand biological function with extraordinary precision, and they have given pharmacologists and physicians a significant fraction of modern therapeutic agents. Naturally occurring small molecules, or their derivatives, are significant contributors to new drugs (60% of new cancer drugs and 75% of new anti-infectives from 1981 to 2002) and are well represented among the top-selling pharmaceutical agents. A few examples of these remarkable molecules are shown in figure 1.

While biologically active small molecules have been isolated from many sources, the molecules in figure 1 are all produced by microorganisms that live in soil, and small molecules from soil microbes have made the greatest contribution to therapeutic agents. The bacterial realm of life has much greater genetic diversity than the more familiar eukaryotic realm of plants and animals. For example, two common bacteria, *Eschericia coli* and *Bacillus subtilis*, are genetically more distant from each other than humans are from corn. Different groups of bacteria produce different types of molecules. The actinomycetes, a group of gram-positive bacteria characterized by branching filaments that resemble the mycelia of filamentous fungi, produce rapamycin, staurosporine, and dynemicin, among many others. Rapamycin, which was first found from a microbe in a soil sample collected on Easter Island, was discovered

---

<sup>1</sup>Read 28 April 2005, as part of the symposium "Discovery and Invention in Contemporary Chemistry."

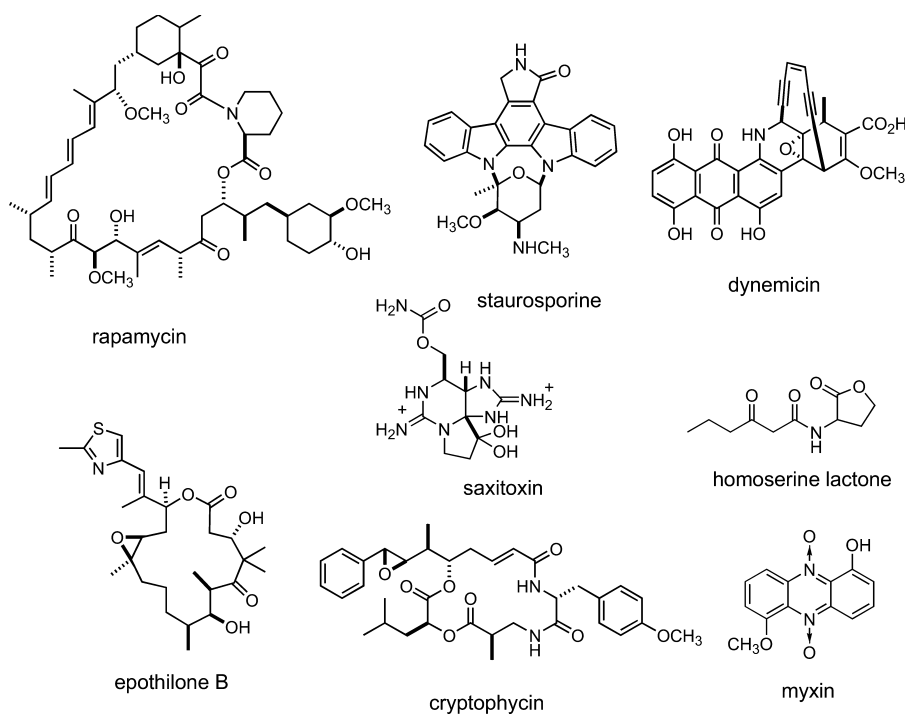


FIGURE 1. Compounds produced by cultured soil bacteria. The top row (rapamycin, staurosporine, and dynemicin) were all produced by actinomycetes; epothilone was produced by a myxobacteria, saxitoxin and cryptophycin were produced by cyanobacteria; and homoserine lactone and myxin were produced by Proteobacteria.

because of its antifungal activity. Today it is used to prevent the rejection of transplanted organs and is being studied as a treatment for cancers.

The other molecules in figure 1 have similar histories and possible uses. Staurosporine was the first compound to show that partially selective kinase inhibitors could be discovered (or invented), and while its promiscuity for various kinases limited its clinical utility, it paved the way for the recently introduced anticancer agents such as Gleevec and Iressa. Dynemicin is one of the “enediynes” class of antitumor antibiotics, and its mode of action—a spectacular molecular rearrangement that yields a potent DNA-damaging agent—is found in a small family of very cytotoxic compounds. Versions of enediynes that are targeted to home in on cancer cells are currently in clinical trials as anticancer agents. A derivative of epothilone B, which comes from a myxobacteria or slime mold, is also in clinical trials as an anticancer agent. Saxitoxin and cryptophycin, which are both produced by cyanobacteria (blue-green algae), are used to probe ion channels in nerve cells and as a possible

anticancer agent, respectively. The last two, myxin and a homoserine lactone, are from the large class of Proteobacteria. Homoserine lactone is a signaling molecule that bacteria use to measure their cell density, and myxin is a broad-spectrum antibiotic. These are just a few of the many valuable molecules produced by soil-dwelling bacteria.

Ironically, while we know a lot about these molecules and their actions in mammalian systems, we know much less about their natural role. Why, for example, does a soil-dwelling microbe on Easter Island make a molecule with an exquisite ability to disrupt human cell signaling? The usual answer, based more on belief than experiment, is that such molecules serve a defensive (or offensive) function and contribute to the survival of the producing species. Rapamycin prevents the growth of fungi by altering a fungal signaling pathway to trigger a starvation response. As the fungi hunker down, metabolically speaking, to await better times, the bacterial producer has one less competitor for nutrients. Signaling pathways are remarkably conserved, and human and fungal signaling pathways are similar enough in design, although different in purpose, that rapamycin can also switch human cellular pathways into arresting cell division. Finally, these molecules are objects of great esthetic beauty, at least to trained observers. Their structures often reveal familiar molecular motifs in strikingly unfamiliar combinations. Evolution reuses and recombines standard molecular features to produce new functionality. For example, there is nothing unusual about double and triple bonds between carbon atoms, but when they are combined in the way they are in dynemicin and the other enediynes, the combination creates a distinctly new type of molecule with amazing reactivity.

It is not surprising that microbes are such prolific creators of small molecules. First, microbes are wonderful chemists and have a primary and secondary metabolic repertoire that dwarfs that of humans and other eukaryotes. Second, microbes interact with the world through chemical messages carried by small molecules. Small-molecule signaling systems are the eyes and ears by which they perceive their environment and react to environmental changes. For example, the homoserine lactones are used to coordinate group activity in bacteria, and pathogenic bacteria will wait until they have a substantial population, measured by the homoserine lactone level, before revealing their pathogenic nature. The production of small molecules is tightly regulated, and metabolically expensive small molecules are produced only when needed. Often the signal to produce a small molecule is another small molecule. An emphasis on defensive (or offensive) compounds is undoubtedly why microbial small molecules so frequently distinguish themselves as antibiotics and anticancer agents. Third, the bacteria have had a long

evolutionary history. They are the direct descendants of the earliest life forms on the planet, and they have a short generation time.

#### COMPOUNDS FROM CULTURED BACTERIA

The search for biologically active small molecules produced by soil microbes is conceptually simple. Soil samples are collected and microbes are cultured on agar plates containing the necessary nutrients. Culturing microbes, identifying strains whose extracts have interesting biological activities, and purifying the molecules responsible for the activities, led to the discovery of all of the compounds in figure 1. This protocol for discovering compounds is not new, and its place in the popular imagination was undoubtedly secured by the accounts of Fleming's discovery of a penicillin-producing fungus that was killing bacteria on a contaminated culture plate.

Shortly after Selman Waksman's seminal discoveries of actinomycin and streptothricin and Albert Schatz and Elizabeth Bugie's discovery of streptomycin, the search for biologically active small molecules, especially antibiotics, from cultured soil bacteria became a thriving industry. Pharmaceutical companies literally scoured the earth and amassed large collections of soil and soil-derived microbes. Those are the collections that gave rise to most of our antibiotics, most of our anticancer agents, and many of our other drugs. Today, sixty years later, many large pharmaceutical companies have downsized or completely eliminated these natural products discovery programs because of their failure to discover new molecules. Most of the antibiotics used today were discovered between 1940 and 1970, and the search for new ones since that time has had limited success. One widely cited estimate states that the chances that a molecule discovered from a microbe will be new is less than 1 in 500. Do these discouraging statistics reflect an exhausted source or the limitations of the approach just described? Many scientists believe that the problem lies in the approach, not in the source.

#### COMPOUNDS FROM UNCULTURED BACTERIA

Today it is clear that the potential of soil (and other) microbes to produce interesting small molecules has certainly not been exhausted, or at least that the number of bacteria to be studied has certainly not been exhausted. There are two kinds of bacteria in soil: those that can be cultured under known laboratory conditions and those that cannot. This distinction has been known since the 1930s, when scientists reported that they could see bacteria in soil samples that never showed up on the culture plates. By the late 1990s, the magnitude of this "plate

count anomaly” became clear, and the startling result was that fewer than 1%, perhaps fewer than 0.1%, of the bacteria in soil could be grown using known culture conditions. The discrepancy isn’t limited to soil. Other environments, including the human outer ear, mouth, and atherosclerotic plaque, also showed large numbers of previously uncultured bacteria. How were those discrepancies discovered? All living things on this planet store their genetic information in DNA, transcribe that information into RNA, and translate the RNA-based information into proteins, which are used to build cells and carry out metabolic processes. A molecular machine called the ribosome translates RNA into proteins. Ribosomal genes have been intensively studied, and one fragment in particular, the 16S rDNA fragment, has been used to trace the course of evolution and the relatedness of organisms. DNA can be extracted directly from environmental samples, and the polymerase chain reaction (PCR) can be used to generate multiple copies of the 16S rDNA sequences. Analysis of these sequences showed that essentially all of the 16S rDNA obtained from soil in this “culture independent” method was not present in the existing databases from cultured organisms. This imbalance can be explained only if the cultured organisms represent a tiny fraction of the total organisms in a sample. Taken at face value, these results argue that all of the interesting compounds isolated from soil bacteria are a tiny fraction of the possible compounds that could be isolated. There could be a hundred to a thousand times as many compounds that have not yet been discovered, but discovering compounds produced by organisms that cannot be cultured represents a significant obstacle. Many laboratories have taken up the challenge of discovering these molecules, and my laboratory has used the approach outlined below.

The bacteria present in a soil (or any environmental sample) are broken open by heating in detergent (fig. 2). The freed DNA is collected and purified using standard techniques. The DNA is then put into an easily cultured bacterial host such as *Escherichia coli*, and clones making antibiotics are discovered through a simple overlay assay. This approach assumes that the genes that govern the production of small molecules in bacteria are on a continuous stretch of DNA. For characterized pathways—the genes that produce the proteins that make the small molecule, that confer resistance to the small molecule, and that regulate the production of the small molecule—are all on a continuous stretch of DNA, so even if this genetic organization is not universally followed, it seems like a good bet for an initial search. Reliance on the clustering of genes is not based only on historical precedent. A compact organization of the genes for small molecule production will be followed because only by keeping all the genes together can their collective utility confer a survival benefit to their possessor. Imagine, for example, the

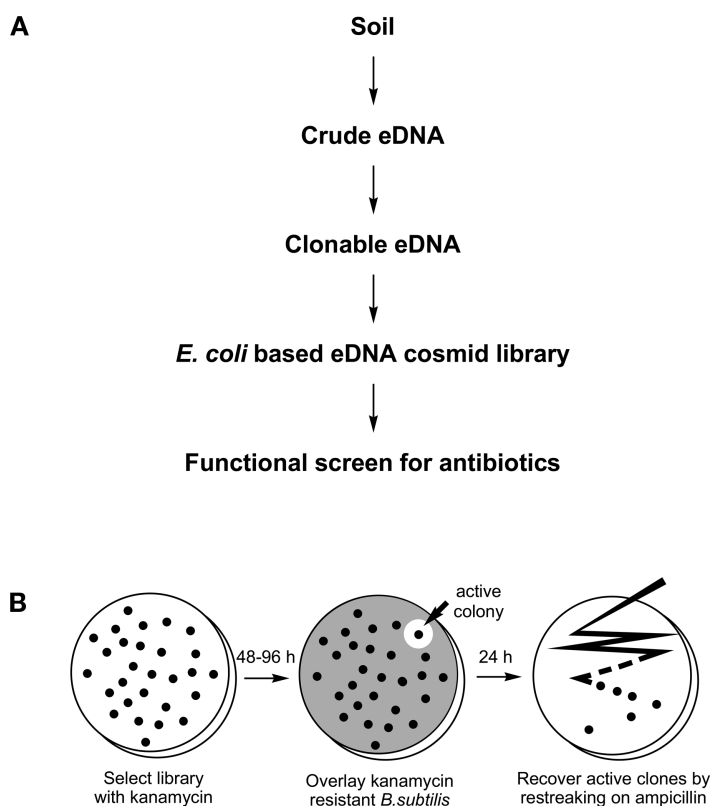


FIGURE 2. A. The general procedure for isolating environmental DNA from soil. B. The overlay assay for antibiotic discovery.

plight of a bacterium that gets a copy of the genes for making an antibiotic without also getting the resistance gene that will keep it from getting killed by the antibiotic. Many of these small molecule biosynthetic fragments of DNA have been moved around in cultured bacteria, and putting these fragments into an appropriate host confers the ability to make the small molecule. For example, the genes for the beta-lactam antibiotic cephalosporin C have been moved from the original producing bacteria (*Lysobacter lactamgenus*) and put in the alternative host (*Pseudomonas putida*). The transformed *P. putida* makes cephalosporin C, an ability it did not possess with its initial genetic endowment.

#### SCREEN BY FUNCTION

To see whether this plan of capturing pathways was feasible, DNA was extracted directly from soil, cleaned up, sorted by size, and inserted into *E. coli*, an easily cultured bacteria with well-characterized genetics. In

this fashion libraries containing up to several million *E. coli* clones could be prepared, although we typically make much smaller libraries. The demanding job was not preparing the libraries, but discovering those clones making interesting small molecules. The odds of capturing a complete pathway that could be expressed in *E. coli* were small, conceivably zero, so a quick way to discover the few interesting library members was developed. A thin layer of agar containing *B. subtilis* was carefully layered on top of the *E. coli* colonies, and clones producing an antibiotic that killed *B. subtilis* could be identified by a zone of inhibition—a clear zone where the *B. subtilis* didn't grow (fig. 2B). The laboratory has used this method to screen several million colonies, and many interesting small molecules have been identified.

One of the compounds will be explored in more detail (fig. 3). A library prepared from soil collected in Boston had a colony with a faint orange cast and a large zone of inhibition—two indications that it was making a small molecule antibiotic. Isolation and characterization of the small molecule, which is relatively unstable, revealed an isonitrile derived from the amino acid tryptophan (fig. 3). Natural products containing the isonitrile functionality are known, but they are relatively rare and have posed two questions that have been unanswered for the last fifty years: 1) where does the carbon atom of the isonitrile come from? and 2) how is the triple bond between the carbon and nitrogen made? Knowing the biosynthetic genes responsible for isonitrile production would be a powerful start to answering such questions, but no laboratories had succeeded in identifying them. With the DNA-based procedure just described, finding the responsible genes is easy, and only two genes, which were called *isnA* and *isnB*, were required. Having the genes expressed in *E. coli* was also a great advantage, since *E. coli* mutants could be used to find the origins of the nitrogen and carbon atoms that make up the isonitrile.

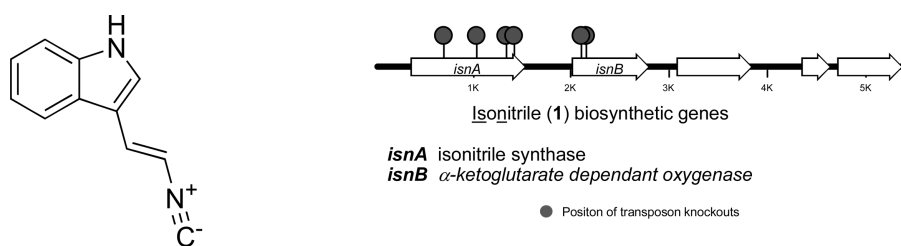


FIGURE 3. *Left:* an isonitrile-containing molecule isolated using a functional antibiotic assay. *Right:* the genes responsible for isonitrile biosynthesis. The lollipop figures represent disruptions in the *isnA* and *isnB* genes that knock out antibiotic production.

## SCREEN BY SEQUENCE

While the procedure described above, capturing DNA from soil and discovering small molecules using a functional assay for antibiotic activity, works, it is likely to miss some of the compounds. The captured DNA can be thought of as a recipe, and to expand the analogy, there are many ways in which a recipe, although perfectly correct, would not be useful. It might be written in a language that could not be understood or require ingredients that were not available. If these shortcomings were translated into the language of microbiology, they would be called “different codon usage” and “wrong metabolic background,” and either would cause the procedure just described to fail. Or the recipe might be for a dessert and the situation demands an appetizer, which in the laboratory procedure would correspond to having the wrong assay. The assay described would find antibiotics, for example, but it would likely miss compounds effective against diabetes. Or the recipe might be for a single serving, and a large family must be fed. In the discovery scheme this would correspond to having too weak a signal to detect—the expression level is just too low. When these shortcomings of the approach are outlined, it might seem surprising that it worked at all, and its success argues that the number of undiscovered compounds is so large, that even a very limited approach will find many of them. But the limitations also underline the need for a more thorough method of discovery.

Rather than searching on function, searching for antibiotic molecules, one could search by sequence, searching for a particular gene that might lead to a small molecule. Searching by sequence (or gene) would get around many of the limitations seen in the functional approach. Once the gene was found, DNA sequencing would reveal if a whole pathway were present, the appropriate type of bacteria to use for expression, and the answers to similar questions. The search, in principle, would identify a small molecule irrespective of its biological activity or potency.

We have recently tried to implement such a sequence-based discovery approach with encouraging results. We wanted to use a relatively rare gene so that we wouldn't have to wade through a large number of hits, but we also wanted to use a gene that we were confident would be associated with small molecule production. The *isnA* gene, the one involved in producing isonitriles, seemed like a good choice. A sequence search can take two forms: it can go through the database of existing genome sequences or through the DNA extracted from soil. We first began to look in the existing genome sequences, and were surprised to find *isnA* relatives in at least half a dozen bacteria. The bacteria were a

rogues' gallery of pathogens, and bacteria causing cholera, complications of cystic fibrosis, a disease of horses, Legionnaires' disease, and a disease of carrots were all found to contain *isnA*. We identified the biosynthetic sequences in the published genomes, cloned them out of their natural hosts, and put them into an *E. coli* background. This procedure was a much easier way to identify the small molecule produced because the *E. coli* strain that we use has a negligible background level of small molecule production, so identifying any new small molecules is relatively straightforward. In every case, a new small molecule was discovered, a molecule that had never previously been reported in the scientific literature.

## CONCLUSION

In one sense, the way to discover new compounds in nature is quite simple. Just look in a place where nobody else has looked. Like most simple advice, that is hard to follow because it lacks an explanation of how it can be enacted. Uncultured bacteria are definitely a place to look for new compounds. The cultured bacteria are wonderful chemists, and the uncultured ones should also be good chemists. The approach of extracting DNA directly from the environment and putting it into alternative hosts and looking for new molecules—either by function or by sequence—seems to be a promising road to follow. It has the advantage of intimately linking the small molecule with the DNA that gives rise to its production, and in that way connects natural products chemistry to the fundamental units of evolution. It allows small molecule chemistry to be put into an evolutionary context, which, as Dobzhansky observed many years ago, is the only way to make sense of biology.

## ACKNOWLEDGMENTS

Work in my laboratory is supported by grants from the National Cancer Institute and the Ellison Medical Foundation. I gratefully acknowledge the work by various laboratory members, especially Sean Brady.

## REFERENCES

- Brady, S. F., and J. Clardy. 2000. Long-chain N-acyl amino acid antibiotics isolated from heterologously expressed environmental DNA. *J Am Chem Soc* 122:12903–04.
- . 2002. New natural product families from an environmental DNA (eDNA) gene cluster. *J Am Chem Soc* 124:9968–69.
- . 2004. Long-chain N-acyltyrosine synthases from environmental DNA. *Appl Environ Micro* 70:6865–70.

- Butler, M. S. 2004. The role of natural product chemistry in drug discovery. *J Nat Prod* 67:2141–53.
- Clardy, J., and C. Walsh. 2004. Lessons from natural molecules. *Nature* 432:829–37.
- Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180: 4765–74.
- Newman, D. J., G. M. Cragg, and K. M. Snader. 2003. Natural products as sources of new drugs over the period 1983–2002. *J Nat Prod* 66:1022–37.
- Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. *Science* 276:734–40.
- Stone, M. J., and D. H. Williams. 1992. On the evolution of functional secondary metabolites (natural products). *Mol Microbiol* 6:29–34.