

ADVERTISEMENT



Applying Novel Technology in Microwave Synthesis

CEM

DRUG DISCOVERY

How DNA-encoded libraries are revolutionizing drug discovery

With the bar-coding technology, drugmakers leverage the chemistry of large numbers

by **Bethany Halford**

JUNE 19, 2017 | APPEARED IN **VOLUME 95, ISSUE 25**



ADVERTISEMENT

ACS
ES&T | Water

EDITOR-IN-CHIEF
Professor Shane A. Snyder
Nanyang Technological University, Singapore

Open for Submissions

MOST POPULAR IN PHARMACEUTICALS

**Pfizer's novel COVID-19
antiviral heads to clinical trials**

**Adenoviral vectors are the new
COVID-19 vaccine front-
runners. Can they overcome
their checkered past?**

Can Pepcid treat COVID-19?

**Gilead's \$21 billion purchase
of Immunomedics boosts ADC
field**

**CRISPR start-up Graphite Bio
launches with \$45 million for
targeted DNA integration**

**How Does Acetaminophen
Work? Researchers Still Aren't
Sure**



Credit: C&EN

Forty trillion is the kind of number that gives one pause. Consider it written out with its 13 zeros: 40,000,000,000,000. Assembling and maintaining a collection of 40 trillion of anything seems like a mind-bogglingly massive task. But in February the Danish

biopharmaceutical company **Nuevolution** announced that it had created a library of 40 trillion unique molecules—quite possibly the largest collection of synthetic compounds in the world.

You might think it would require every building in Copenhagen to store batches of 40 trillion different compounds. Not so, says Alex Haahr Gouliaev, Nuevolution's chief executive officer. "All of that fits into an Eppendorf tube and is handled by one person for screening," he says.

The substance that makes it possible to maintain this multitudinous mixture of molecules is the same substance that contains the code of life—DNA. Nuevolution covalently attaches a short, unique strand of DNA to each of its 40 trillion compounds. Instead of holding the directions for life, though, these DNA strands encode the recipe used to synthesize each linked molecule. This trick enables the firm to store all the compounds as a mixture in a small volume and later sequence, or read, them out. As the cost for DNA sequencing plummets and the repertoire of DNA-compatible chemical reactions grows, these so-called DNA-encoded libraries are becoming a go-to resource for finding new drug candidates and research tools for large pharmaceutical companies, small biotechs, and academics alike.

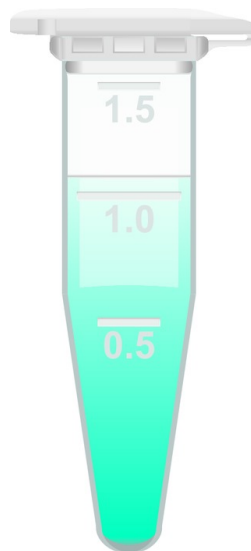
"DNA-encoded libraries are revolutionary," says **Roger D. Kornberg**, a biochemist at Stanford University School of Medicine and winner of the 2006 Nobel Prize in Chemistry. "I think they represent the most innovative and broadly significant advance in chemistry in the past decade or more. Some of my chemical colleagues who develop beautiful new chemistry might be offended by the breadth of that remark, but suffice it to say, this is a major advance."

A dizzying number of deals in the DNA-encoded library space over the past year demonstrate the pharmaceutical industry's growing excitement over the technology. Last October, **Amgen** and Nuevolution inked a collaboration for the former to use the latter's DNA-encoded libraries to search for drug candidates against multiple targets in oncology and neuroscience. GlaxoSmithKline, a world leader in DNA-encoded library technologies, established a partnership with **Warp Drive Bio** in March to create a library aimed at targets previously considered "undruggable." **HitGen**, a Chinese company that specializes in DNA-encoded libraries, has set up partnerships with **Johnson & Johnson**, **Merck & Co.**, **Pfizer**, and the **California Institute for Biomedical Research** over the past nine months. And just last month **X-Chem Pharmaceuticals**, another company that specializes in DNA-encoded libraries, announced it would be collaborating with Vertex Pharmaceuticals.

Companies are also expanding their in-house efforts with DNA-encoded libraries. In February, Novartis announced that it would use the technology to ramp up its compound collection from 3 million molecules to 300 million over the next three years.

IN BRIEF

DNA-encoded libraries let researchers screen millions, billions, and even trillions of chemical compounds in a single, simple experiment, thanks to a DNA tag that encodes how each component in the library was made. Although the technology was invented 25 years ago, it's only within the past five years that it's become a mainstay of drug discovery. Read on to learn about how the technology works and to read some recent success stories that pharmaceutical companies, biotechs, and academics have achieved using the technology.



Credit: Shutterstock

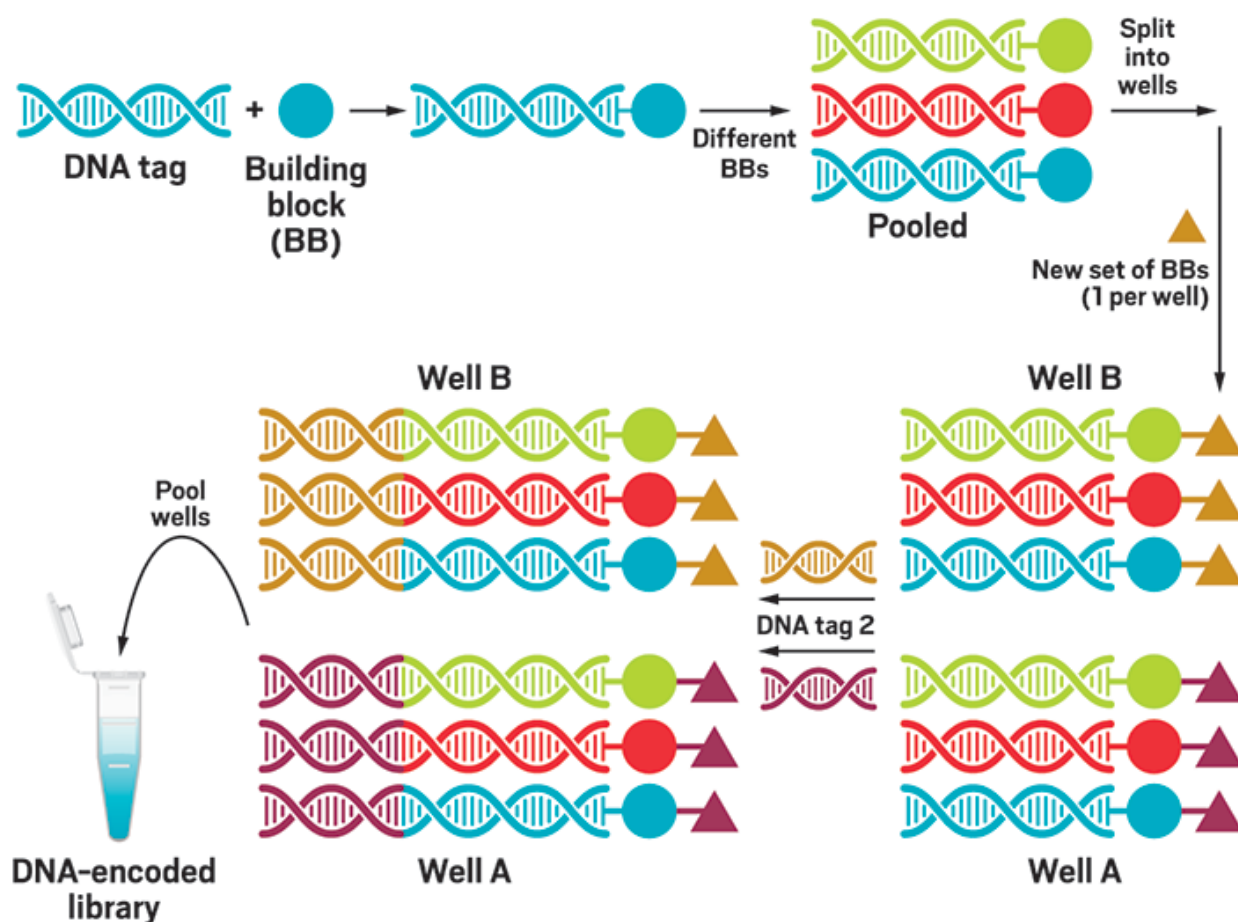
The reason for all this activity is obvious, Kornberg says. The standard for testing compounds in the pharmaceutical industry has for a long time been the high-throughput screen, in which scientists interrogate a library of a couple million compounds one by one to see if they affect the function of a target of interest. “To do all of that costs on the order of a billion dollars and requires instrumentation that occupies space the size of the building I am sitting in at the moment,” Kornberg says from his office at Stanford’s three-story Fairchild building.

By comparison, a DNA-encoded library of billions or even trillions of compounds can fit into a space the size of an Eppendorf tube and costs just tens to hundreds of thousands of dollars to create and use. That’s because the DNA-encoded library is made, stored, and screened as a mixture.

“Since we can screen them as a mixture, there’s really no limit to the number of molecules we can put into the mixture,” explains Matthew A. Clark, senior vice president of research at X-Chem.

Building a library

The most popular method for creating a DNA-encoded library involves assembling DNA and a small organic building block. These components are split into wells, and another building block is added, along with a second piece of identifying DNA. By repeating this process with large numbers of building blocks, it is possible to create large libraries quickly.



Credit: C&EN/Adapted from *The Scientist*

Constructing and reading the library

Although scientists can use a few different methods to make a DNA-encoded library, the one they use most often treats the DNA like a bar code. They start by attaching a short piece of DNA to a small organic functional group—an aliphatic amine, for example. That basic building block is then split into wells in a plate, where it undergoes a chemical reaction with a different building block in each well. Then researchers add a unique bit of DNA, anywhere from seven to 15 base pairs long, to each well and connect, or ligate, it to the existing DNA, creating a code for the reaction that just took place. The contents of all the wells are then pooled and split up again. The process is repeated. In this manner, it's possible to build a library of considerable size in just a few iterations.

To screen a DNA-encoded library, researchers combine the mix of compounds with a biological target such as an enzyme. Anything that doesn't bind to the target washes away. The scientists then denature the target, collect the resulting batch of hits, and incubate them with a fresh target to ensure the best binders remain. This process gets repeated for a third time. Only vanishingly small amounts of the compounds that bind the target remain after these repeated screenings, so to determine their identities, the DNA on each compound must be amplified and sequenced. By analyzing the sequences, scientists can read the DNA bar code and tell which reactions and building blocks were used to make the compounds that bind best. Chemists then resynthesize those compounds without the DNA tag and test them with the target again to see if they have any biological effects.

“DNA is quite robust—after all, we are still digging out DNA from dinosaurs—but it is still fragile with respect to pH and temperature.

— *Frédéric Berst, scientist, Novartis*

“DNA-encoded libraries are a rejuvenation of the combinatorial chemistry concepts of the 1990s propelled into the 21st century,” says Frédéric Berst, a scientist who works on DNA-encoded libraries at Novartis. “You can deeply and routinely sample huge chemical collections in a comparatively easy-to-run experiment.”

“If you have 3 million compounds, to screen them all with high-throughput screening is really a lot of work,” says Robert A. Goodnow Jr., a scientist with Pharmaron and editor of **“A Handbook for DNA-Encoded Chemistry: Theory and Applications for Exploring Chemical Space and Drug Discovery.”** But with DNA-encoded chemistry, you can put hundreds of millions, billions, or even trillions of compounds in front of a target. “You simply could not assay a billion compounds in a high-throughput screening format,” Goodnow says. “It’s just not possible in terms of time and money.”



SPONSORED CONTENT

Applying Novel Technology in Microwave Synthesis

by CEM Corporation

Besides the leap in the number of compounds that can be screened in a single experiment, the technology offers an additional advantage. It's possible to do many screenings in parallel with DNA-encoded libraries, says Johannes Ottl, another Novartis scientist who works with the technology. That can't be said for high-throughput screening.

For example, it's relatively easy to find kinase inhibitors but challenging to find inhibitors that are specific for a particular kinase. If you wanted to test a high-throughput screening collection of 1 million compounds against 50 specific kinases, you'd need to conduct 50 million experiments. To do the same type of screening with a DNA-encoded library would take only 50 experiments and could potentially identify compounds that bind to a specific kinase.

"We don't want to make it sound too simple because there is a lot of due diligence you need to do to run such a project," Ottl says. "But the up-front work—the assay development—is quite simple compared to many other approaches in lead finding." Basically, he says, scientists are just fishing for binders and don't need to create an assay that measures a target's biological function.

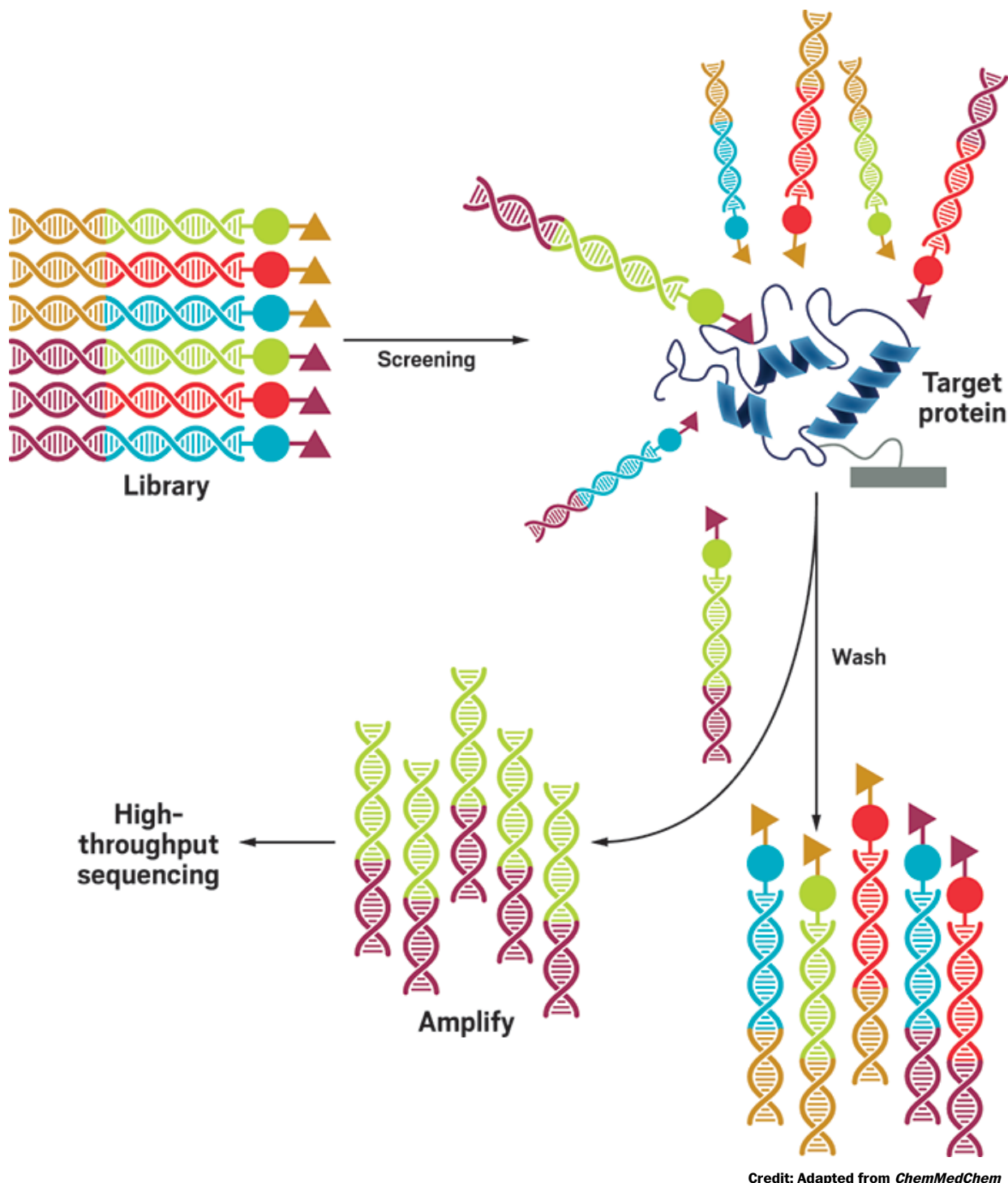
A little history

The concept of DNA-encoded libraries was introduced 25 years ago by **Richard Lerner**, a chemist at Scripps Research Institute California, and his colleague **Sydney Brenner**, cowinner of the 2002 Nobel Prize in Physiology or Medicine. The pair published a paper that's often described as a "thought experiment" (*Proc. Natl. Acad. Sci. USA* 1992, DOI: **10.1073/pnas.89.12.5381**). They also put their pipettes into action to make a small DNA-encoded library and patented the idea around the same time (**U.S. Patent No. 5573905**).

Lerner recalls that the two came up with the concept when discussing the difference between chemistry and biology. They reasoned that small molecules, such as drugs and natural products, differ from biological molecules in that they do not carry information in the form of a code. "They don't tell you who they are," Lerner explains, "and secondly, they don't replicate." Lerner and Brenner reckoned that they could give molecules a replicable identity by putting a piece of DNA on them after each step in a chemical synthesis.

Library screening

Screening a DNA-encoded library involves exposing a biological target of interest to that library. Components that don't bind are washed away, while those that do are amplified and sequenced.



Typically, Lerner says, “large numbers are the enemy of identification in organic chemistry.” But with DNA-encoded library technology, scientists can take large numbers of molecules and give each one an identifying marker that carries information, Lerner points out. That information can be replicated, he says, adding, “It’s hard to beat that sort of power.”



The idea, however, languished for at least a decade.

“For a long time I think the technology was not readily available for people to try, nor did they understand it well enough to say ‘I want to apply it,’ ” Goodnow says. But, he adds, that attitude has changed in the past five years.

The LC Side of "Do More with Less"

by Agilent Technologies

"People have become much more aware that this presents a real opportunity to find hits. It's another tool in the toolbox."

"It's remarkable that it works so well," says Barry A. Morgan, HitGen's chief scientific officer. "The reason that it works is really a tribute to the fundamental

developments over the last 30 years in our ability to manipulate and sequence DNA."

In the early 2000s, Morgan worked for Praecis Pharmaceuticals (which was acquired by GSK in 2007), one of the first companies to explore DNA-encoded libraries. When the firm started working on the technology, he recalls, current high-throughput DNA-sequencing methods weren't available. But about six months into the project, he and his Praecis colleagues found a company called 454 Life Sciences that had sequencing methods perfectly suited to DNA-encoded libraries.

"We wouldn't be able to make such large libraries and deconvolute them if the current sequencing methods were not available," Morgan says.

Gouliaev says that when Nuevolution was getting started in the early 2000s, pharma companies and venture capitalists would tell him that it didn't make sense to synthesize such big libraries. They were put off by previous efforts in combinatorial chemistry, wherein chemists prepared tens of thousands to millions of small molecules as a mixture and screened them for useful properties. "They would say, 'Don't you know combinatorial chemistry failed?' And, 'Having DNA will limit your chemistry so much that you can't make the molecules we'd be interested in.'"

"We needed to prove ourselves," Gouliaev continues, "and it took us quite a few years to get something that was robust, reliable, and would have high diversity of truly druglike small molecules."

X-Chem's Clark agrees that many were skeptical about DNA-encoded libraries because of the failure of combinatorial chemistry in the 1990s. "The best way to overcome skepticism is with data, and there's been enough data reported in the last five years that it would be very difficult to maintain that sort of skepticism," Clark says.

Success stories

Several DNA-encoded library success stories have emerged just this year. **GSK** advanced its compound GSK2982772—which came about from DNA-encoded library work—to Phase IIa clinical trials in patients with psoriasis, rheumatoid arthritis, and ulcerative colitis. GSK2982772 inhibits receptor interacting protein 1 kinase, or RIP1 kinase, an enzyme that's been linked to inflammation.

Looking to develop an inhibitor for RIP1 kinase, scientists at GSK first screened the company's set of known kinase inhibitors, but they were unable to find molecules that had the druglike properties they were looking for, and they also found that hits from this set of compounds weren't selective for RIP1; they inhibited other kinases as well.

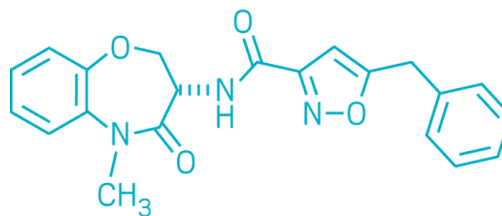
They also screened GSK's high-throughput collection of roughly 2 million compounds and identified a RIP1 inhibitor, but that compound had challenges. In particular, it didn't get into the bloodstream of rodents when given orally. By far the most promising lead, a compound known as GSK^ˆ481, was obtained by screening a DNA-encoded library of approximately 7.7 billion compounds against RIP1 kinase. GSK^ˆ481 turned

out to be extremely potent as well as highly specific to RIP1 kinase (*J. Med. Chem.* 2016, DOI:

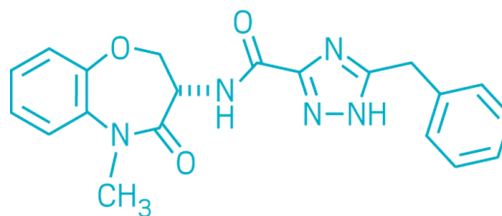
10.1021/acs.jmedchem.5b01898).

But the scientists thought they could improve GSK´ 481’s pharmacokinetics. Using a traditional medicinal chemistry approach, they eventually wound up swapping GSK´ 481’s isoxazole for a triazole to get their clinical candidate GSK2982772 (*J. Med. Chem.* 2017, DOI: 10.1021/acs.jmedchem.6b01751).

“On paper it looks like it was just a tweak in a few atoms,” says Christopher P. Davie, manager of discovery chemistry at GSK who leads its efforts on DNA-compatible reaction development and encoded library synthesis. “But a ton of medicinal chemistry work went into it.”



GSK´ 481



GSK2982772

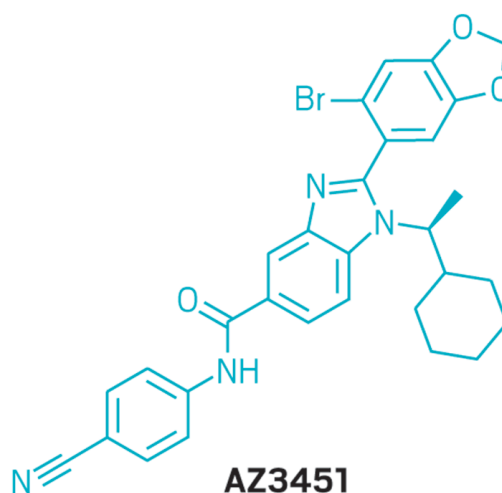
ADVERTISEMENT



In another recent success story, just last month, researchers at AstraZeneca, Heptares Therapeutics, and X-Chem published the crystal structure of two allosteric ligands

bound to a G protein-coupled receptor (GPCR) called protease-activated receptor 2, or PAR2. One of those allosteric ligands—AZ3451—was identified using DNA-encoded libraries from X-Chem (*Nature* 2017, DOI:

10.1038/nature22309).



AZ3451

PAR2 has been implicated in a wide range of diseases, including cancer and inflammation. Allosteric binders of this target could prevent the structural rearrangements PAR2 needs to undergo to become active and participate in signaling. The researchers hope AZ3451 will help guide them in the development of selective PAR2 antagonists for a range of therapeutic uses.

This year’s success stories aren’t just limited to the pharmaceutical industry. “Every academic who’s a biologist and has a target would like to do chemistry, but they have no access to chemical matter,” Scripps’s Lerner points out. DNA-encoded libraries now make it possible for academics to access compounds that pharmaceutical companies struggled for many years to develop, he says.

One recent academic success comes from 2012 Chemistry Nobel Laureate **Robert J. Lefkowitz’s** lab at Duke University. Lefkowitz’s team used a 190 million-compound DNA-encoded library from Nuevolution to find an allosteric modulator for the β 2-adrenergic receptor, another GPCR. Screening for ligands of GPCRs has, in the past, been a cumbersome and labor-intensive process. The method developed by the Lefkowitz lab using DNA-encoded libraries is broadly applicable, the researchers note, and could potentially lead to more therapeutics that target these receptors (*Proc. Natl. Acad. Sci. USA* 2017, DOI: **10.1073/pnas.1620645114**).

Along with the recent success stories, many DNA-encoded library makers point to the economic advantage their technology provides. “You can make from scratch DNA-encoded libraries for a relatively small investment compared with accruing a high-throughput screening collection,” HitGen’s Morgan points out. “And then you can interrogate those libraries very efficiently and effectively within a period of a few weeks.”

“With DNA-encoded chemistry, because you’re making such large numbers of compounds in very small quantities, the cost of production of that mixture of compounds is orders of magnitude smaller” than previous methods, says Pharmaron’s Goodnow, who broke down the cost savings in a recent paper in *Nature Reviews Drug Discovery* (2016, DOI: **10.1038/nrd.2016.213**).



SPONSORED CONTENT

From lab to pilot and production scale

by Pope Scientific

To create and interrogate a conventional high-throughput screening collection of 1 million compounds costs between \$400 million and \$2 billion, roughly \$1,100 per compound, by Goodnow’s estimate. A DNA-encoded library of 800 million compounds, on the other hand, costs about \$150,000 for materials to create and screen—approximately \$0.0002 per compound.

This makes DNA-encoded libraries a good starting point for small companies, start-ups, or academics who don’t already have a large high-throughput screening collection at their disposal, Novartis’s Ottl says. “You don’t need to invest much in automation, and you don’t need to invest a lot in compounds up front.”

But even with the cost savings and massive expansion of chemical space to explore, scientists who work with DNA-encoded libraries say the technology’s not a panacea. “It’s a complement to other existing methods. It’s not better than high-throughput screening,” Goodnow says. “It’s just a different way to go about it. Ideally people would want to do both.”

Challenges

Despite their revolutionary stature, DNA-encoded libraries are not without their challenges. Often chemists wonder if the large DNA bar code attached to a compound will interfere with how it binds to a target. Ideally, the DNA tag would face away from where the compound is binding to a target, explains GSK’s Davie. That’s always observed in crystal structures of library ligands that have successfully bound to their target. Of course, that’s not always going to happen for unsuccessful ligands, he says, but while the DNA tag is a natural constraint of DNA-encoded library technology, it’s not a showstopper.

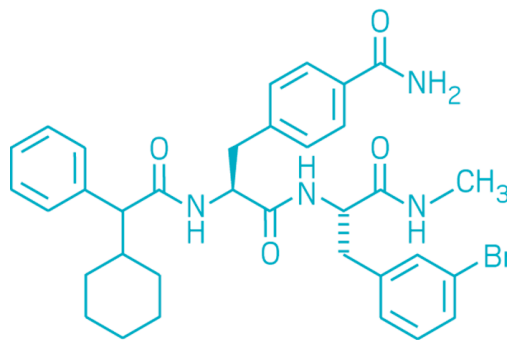
Another constraint is that any chemistry used to construct a DNA-encoded library must be able to tolerate water because DNA requires an aqueous solution. The reaction conditions also have to keep the DNA intact; damaged DNA can’t be amplified or sequenced.

“DNA is quite robust—after all, we are still digging out DNA from dinosaurs—but it is still fragile with respect to pH and temperature,” Novartis’s Berst says. “Heating DNA in xylenes at 200 °C in the presence of a metal catalyst is not something I would recommend to a budding DNA-encoded library chemist,” he jokes.

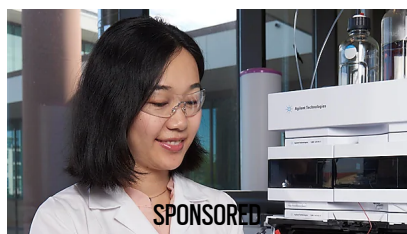
But this limitation also makes developing new DNA-compatible reactions exciting for synthetic chemists. “You put that kind of challenge before a synthetic chemist and it’s like a red cape to a bull,” Scripps’s Lerner says.

GSK’s Davie says that his group has managed to construct components of DNA-encoded libraries in solutions that contain just 20% water. They’ve done ring-closing metathesis reactions as well as cross-couplings, he says, although he admits sometimes they have to use large amounts of catalysts. “The reactions aren’t particularly elegant, but they work,” he says.

Davie thinks the real bottleneck when it comes to DNA-encoded library chemistry comes after library synthesis and screening. He points out that when screening a DNA-encoded library, you turn up only compounds that bind a target. But that’s no guarantee that those hits will have the activity you’re looking for.



**β 2-adrenergic receptor
allosteric modulator**



SPONSORED
CONTENT

The LC Side of "Do More with Less"

by Agilent Technologies

And then there’s the matter of deciding which hits to resynthesize. “There are always more compounds to make than we have resources to make, so we have to prioritize them,” Davie says. Right now, he says, about half the molecules they choose to resynthesize don’t show any activity when screened against their target.

Also, Berst notes, while you might need only 5 mg of a building block when constructing a DNA-encoded library, you will need more than that if you need to resynthesize that hit. When rare building blocks are in hits it can set back the timeline of a resynthesis.

Despite these challenges, DNA-encoded libraries are gaining ground in drug discovery, and scientists see them becoming even more integral to research efforts in the future. “It is truly one of the most profound and original ideas in chemistry, and the consequences are only just beginning to be felt,” Stanford’s Kornberg says. “Once a leap forward in technology takes place, then people begin to think of all kinds of ingenious ways of putting it to use. We’re just at the beginning.”

Sign up for C&EN's must-read weekly newsletter

Email Address

Subscribe »

[Contact us](#) to opt out anytime

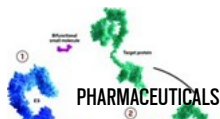
Chemical & Engineering News

ISSN 0009-2347

Copyright © 2020 American Chemical Society

YOU MIGHT ALSO LIKE...

BUSINESS

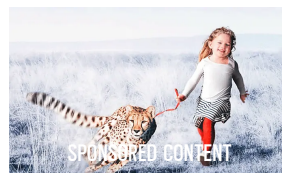


DiCE debuts with major Sanofi pact

Targeted protein degraders are redefining how small molecules look and act

PHARMACEUTICALS

No Strings Attached In Compound Library Deal



Digest different sample types quickly in the same run

by Anton Paar

COMMENTS

Chris Davie

(June 20, 2017 9:48 PM)

If you're interested in learning more, consider attending the (free) 2017 Second Boston Symposium of Encoded Library Platforms (BSELP) at the Brandeis University campus in Waltham, MA on Friday, August 4, 2017

<https://bselp.wordpress.com/agenda/>

Reply »

LEAVE A COMMENT

Name

Email*

Comments by first-time contributors will be reviewed prior to appearing on the site. This review is done by humans and not always immediately. You may be laudatory or critical, but please stay on topic and be respectful of the author and your fellow readers. We reserve the right to remove any comments that are profane, obscene, abusive, or otherwise inappropriate. Email addresses are required so that we can verify you are not a robot overlord and in case we need to contact you about your comment privately. They will not appear on the site.

Submit

*Required to comment

