

nature

DNA Tags Help the Hunt for Drugs

Drug discovery is a daunting process that requires chemists to sift through millions of chemicals to find a single hit. DNA technology can dramatically speed up the search.

- By [Asher Mullard](#), [Nature magazine](#) on February 22, 2016



Drug discovery often starts with researchers assembling large libraries of chemicals and then testing them against a target protein. *Credit: @iStock*

Nestled in a plastic box, in an ordinary laboratory freezer on the second floor of a concrete building in Waltham, Massachusetts, is a clear test tube that contains a concoction of astronomical proportions. The library frozen within, a collection of chemical compounds owned by London-based pharmaceutical company GlaxoSmithKline (GSK), contains as many as 1 trillion unique DNA-tagged molecules — ten times the number of stars in the Milky Way.

This and other such libraries are helping pharma companies and biotechnology firms to quickly identify candidate drugs that can latch onto the proteins involved in disease, especially those proteins that have proved difficult to target. They enable screening to be performed much more rapidly and cheaply than with conventional methods. And academic scientists can also use them to probe basic biology questions and investigate enzymes, receptors and cellular pathways.

Drug discovery often starts with researchers assembling large libraries of chemicals and then testing them against a target protein. Compounds are added individually to wells that contain the target to see

whether they affect its activity. This approach, known as high-throughput screening (HTS), is automated using robotic equipment to test millions of chemicals, but it's still laborious, expensive and not always successful.

Over the past few years, medicinal chemists have been increasing the odds of finding potentially useful compounds by labelling chemical compounds with bits of barcode-like DNA. These DNA-encoded libraries — which can dwarf conventional small-molecule libraries — offer all sorts of advantages to drug discovery. For a start, rather than testing each compound individually, researchers can put all of the DNA-tagged small molecules into a single mixture and then introduce the target protein. Any compounds that bind with the target can be identified easily thanks to their DNA barcodes.

DNA-encoded libraries were first proposed in 1992, in a thought experiment by molecular biologist Sydney Brenner and chemist Richard Lerner, who were then at the Scripps Research Institute in La Jolla, California. They have been gaining momentum ever since. In 2007, GSK acquired one of the firms that pioneered these libraries, Praecis Pharmaceuticals in Waltham, for \$55 million. The drug firms Novartis and Roche, both in Basel, Switzerland, have started their own in-house DNA-encoded-library programmes. A burgeoning group of biotech companies — including X-Chem in Waltham; Vipergen in Copenhagen; Ensemble Therapeutics in Cambridge, Massachusetts; and Philochem in Zurich, Switzerland — has meanwhile built up a who's who list of industry and academic partners that are eager to use the technology.

“People understand now that this is not a fad,” says Robert Goodnow, executive director of the Chemistry Innovation Center at AstraZeneca in Boston, Massachusetts, which collaborates with X-Chem. “It's for real.”

DNA-encoded libraries will not replace HTS: companies have already invested heavily in HTS screening, and there are some compounds that cannot be synthesized using DNA-encoding technologies. Rather, they offer a complementary way to quickly, efficiently and cheaply find chemical structures that bind to new or historically challenging targets, such as ubiquitin ligases, which flag proteins for disposal and could be targeted in cancer therapy.

Big is beautiful

GSK currently has the world's biggest DNA-encoded library: it is an impressive 500,000 times larger than the company's 2-million-compound HTS library.

There are several ways to build DNA-encoded libraries: the biggest ones, like GSK's, are made using an approach called 'DNA recording'. Chemical building blocks, such as amino acids, amines and carboxylic acids, are synthesized and then tagged with a unique DNA barcode through a chemical reaction. A second building block is added to the mix to make a new small molecule, and the DNA barcode is then lengthened. By joining up to four blocks, chemists can create drug-like molecules. And because they have thousands of building blocks to play with, the number of potential combinations is enormous.

Compared with conventional HTS libraries, for which chemists have to test each compound individually, DNA-encoded libraries are easier to maintain and use. A DNA-encoded library can be stored in a single test tube, whereas an HTS library requires robot-filled facilities that are large enough to store each compound individually.

But the true beauty of DNA-encoded libraries, says Chris Arico-Muendel, a manager at GSK in Waltham, is the sheer number of chemical structures it is possible to synthesize. The company's drug-discovery team now uses the DNA-encoded library as frequently, if not more frequently, than the HTS library for new and difficult protein targets. The most advanced compound to emerge from the company's DNA-encoded library so far is GSK2256294, which blocks epoxide hydrolase, an enzyme that is involved in breaking down lipids. This drug candidate came out of GSK's collaboration with Praecis and has completed first-in-human safety studies that may support further evaluation of its use in diabetes, wound healing or as a therapy for chronic obstructive pulmonary disease. "We are pleased with how things are going with DNA-encoded libraries within GSK," says Arico-Muendel.

And as more chemical building blocks are created, along with extra ways to attach them to one another, the breadth of these libraries will continue to expand.

In the near future, DNA-encoded libraries will not only become bigger and broader but might also provide hits that can quickly be moved into the clinic, says X-Chem chief executive Richard Wagner. With conventional screening, medicinal chemists sometimes have to spend many years tweaking compounds to make them specific, potent and safe enough to enter the clinic. "This is just a game of odds," says Wagner. By contrast, the large size of DNA-encoded libraries means that, by chance, some of the compounds they include will be more clinic-ready than others. Although the compounds will still require optimization, "we can get things that are pretty close", he says.

X-Chem, which holds 120 billion compounds in its DNA-encoded library, is already starting to see this in practice. It took just one year to move its most advanced candidate — an autotaxin inhibitor that blocks the conversion of one phospholipid into another — from a screening hit to a clinical candidate. A spin-off company of X-Chem, X-Rx in Wilmington, North Carolina, now plans to start clinical trials of the compound for fibrosis in 2017. Interest in X-Chem's library is spreading across the industry: in the past five years, the company has forged collaborations and licensing agreements with several major pharmaceutical firms — including Roche, AstraZeneca, Bayer, Johnson & Johnson, Pfizer and Sanofi — as well as with a host of biotech and academic partners.

Made to order

Other biotech firms have added an interesting twist. They use the DNA tag not only to identify a compound but also as a template to make it. David Liu, a chemist at Harvard University in Cambridge, and his students developed this 'DNA-templated' approach and used it to build a library of circular molecules called macrocycles. These larger, more-stable, ring-shaped molecules interact with the target at multiple sites, boosting the specificity of the binding reaction. (GSK and X-Chem also have extensive macrocycle collections in their DNA-encoded libraries.)

Liu first creates single-stranded DNA templates that act as guides — these consist of several regions that are complementary to the DNA tags on his chemical building blocks. He then sequentially adds the DNA-tagged building blocks into a reaction vessel, relying on DNA base pairing to physically bring the tagged building blocks close enough together to bind to one another. A final reaction then converts the strings of building blocks into rings, producing macrocycles that are each tethered to a unique DNA barcode.

Constructing a DNA-templated library involves a hefty workload because researchers must design a template for each molecule as well as tagging thousands of building blocks with DNA. As a result, DNA-

templated libraries are smaller than DNA-recorded ones, but they still eclipse HTS libraries in terms of size — and they have other advantages, too. Because scientists know at the outset what compounds they are producing, they can purify the DNA-templated libraries to remove compounds that are tagged inaccurately. This step translates into a high degree of confidence in the hits. By contrast, colossal DNA-recorded libraries may still contain wrongly tagged compounds, and thus might yield hits that will send researchers on wild goose chases.

Liu's 14,000-strong library has already led to a few triumphs. In 2014, his team reported that it had solved a problem that researchers had been struggling with for decades when it found a specific and stable small molecule that can block insulin-degrading enzyme (IDE), which has been linked with type 2 diabetes. He and others have started to unravel the role of IDE in both health and disease, which has led to the identification of other IDE inhibitors. Discussions are under way to develop these into drugs.

Liu has also screened more than 100 other targets, many brought to him by academic collaborators who need small-molecule inhibitors of their pet proteins. “I never would have thought seven years later that this first-generation library would still be providing us with interesting biological discoveries,” he says. “But it has proved to be very fruitful. We have had more hits against targets from our first library than we can follow up on.”

He is nevertheless putting the finishing touches to a second-generation, 256,000-macrocycle library that could open up even more biology. Ensemble Therapeutics, which Liu founded in 2004, now has more than 10 million macrocycles in its library. The company is focusing on targets among the immune checkpoint proteins, which modulate the immune system, and the ubiquitin ligases. It has also granted a license to Novartis to develop one of its finds, a molecule that targets the inflammation-linked protein interleukin-17.

Sweet screens

Once the library is built, the fun of identifying which molecules stick to a target begins. Most researchers rely on 'affinity-based screening' to find those compounds. For this, they engineer the protein target to include a purification tag. They then pass the mixture containing the library and target through a purification column, using the purification tag to pull out the bound pairs. The last step is to read the DNA identifiers linked to the small molecules using a DNA sequencer.

This method can yield results even with minute amounts of a target protein. In one project, remembers Arico-Muendel, academic collaborators wanted to screen an unstable protein that they could produce only in tiny quantities. “They flew it here on dry ice overnight, and we immediately did the entire screen on it,” he says. “And that actually gave some really good hits.” Such experiments are impossible with HTS, because the target protein must be stable and abundant enough to be added into millions of wells before the experiment can begin.

But affinity-based screening has its shortfalls. The clunky DNA tag can sometimes impede interaction with the target, and some potential candidates may be lost. But because the DNA-encoded libraries are so big, screeners are not typically too concerned with these losses. More problematically, small molecules and their tags can bind to the purification column and generate false-positive hits. The purification tag can also interfere with the structure of the target protein, introducing confusion in the data.

Several groups have developed solutions for this problem. Vipergen, a biotech firm that has a DNA-templated library with 50 million molecules, has put its hope in a 'binder trap enrichment' strategy.

Imagine, says Nils Hansen, the company's chief executive, that you could freeze your protein–library mixture and cut it into super-small ice cubes. If the ice cubes are small enough, each will be able to contain only a single target protein. At this size, small molecules that bind to the target will be consistently overrepresented in ice cubes that contain targets, even without a purification strategy. Vipergen has achieved the same effect by putting its screens into water-and-oil emulsions, in which minuscule water droplets stand in for ice cubes. “It's pretty cool,” says Hansen.

Currently, screens of DNA-encoded libraries work best with free-floating, soluble proteins. But many appealing drug targets are embedded in the cell surface, making them impossible to probe with traditional affinity-based screening. For example, an estimated 40% of approved drugs target membrane-bound G-protein-coupled receptors, which sense molecules outside the cell. The technology for screening membrane-bound proteins is evolving, says Goodnow, “but is still kind of a challenge”.

One way forward is to mix a DNA-encoded library with intact cells that overexpress a membrane-bound target. The small molecules can then bind to the targets on the surface of the cell. After the researchers wash away the unbound library, they can identify the bound small-molecule hits by heating up the cells and reading the eluted DNA tags. GSK has used this approach to identify potent inhibitors of a receptor that has been implicated in schizophrenia and disorders of the central nervous system.

X-Chem, too, is starting to see success with screens of membrane-bound proteins. “Historically, the majority of our programmes were on soluble proteins. But there is shift given the recent data we've been able to generate with really difficult membrane-bound proteins,” says Wagner.

With DNA-encoded libraries continuing to expand, and new screening approaches opening up uncharted biological space, he adds, “DNA-encoded libraries are set to become one of the pillars of discovery in the pharmaceutical industry.”