How big should your screening collection be?[†]

Cite this: *Med. Chem. Commun.*,
2018, 9, 1397

Derek Lowe

We're going to need a bigger compound deck...

DOI: 10.1039/c8md90041h

rsc.li/medchemcomm

Here's something that you will never hear said inside a drug discovery group: 'You know, I'm sure that our screening collection is already large and diverse enough. Let's just go with what we have'. At least, I can tell you that I've never heard that one – and if I did, I would immediately back away from the person saying it for fear of what they might do next. All things being equal, an opinion like that would be regarded by many as *prima facie* evidence of mental instability.

Is it? We'll tackle that question in stages. To some extent, it depends on what sort of target you're screening against. For example, many drug companies have collections that are well-stocked with kinase inhibitors – and probably feel that they're a bit too well stocked with them, and wish that the deck had a lower proportion of the things. If you work for such a company and are screening a kinase target, then you should expect at least a few legitimate hits. These may not be selective, or in chemical classes that you want to work in again just yet, but they will be real. And in that case, yes, your screening collection is sufficiently large.

But things will be different if you're working on a more unusual target, particularly one from a whole class of proteins that hasn't been investigated much. Here, it's possible to screen even large compound collections and come

up with nothing that's particularly appealing (or even worse, nothing that turns out to be particularly real). In such cases, people are wont to blame the collection itself – if the darn thing were larger, or if it weren't full of methyl-ethyl-butyl-futile analogues from past projects, or weren't so packed with (say) kinase inhibitors, well, we might be able to make some progress around here for once. There is no purer form of the-grass-must-be-greener longing in early-stage drug discovery than is brought on by thoughts of a better screening collection. They're out there somewhere! Real hits, real leads – why, this project could already be rolling along! But here we sit, becalmed in a wide sea of apparently worthless compounds. Who put all this stuff in the deck, anyway?

These are the unrequited desires that led to DNA-encoded libraries. A good-sized compound collection is in the million-structure range; maybe even a few million, in the more well-stocked parts of the business. But most DNA-encoded libraries probably have ten times that many compounds (or more) in a single little plastic vial: wouldn't you expect a hit if you could screen ten times as many compounds? Admittedly, those compounds will all be derived from the same core structure, but come on! Ten or 20 million of them! And that's just one of the libraries. I guarantee you, no committee charged with expanding the screening deck is planning on making it only ten times bigger;

anyone seriously doing that sort of screening probably has a whole lineup of different ones, easily pushing towards 500 million compounds in total. You don't think that there's a hit in there somewhere?

To be honest, the answer is still 'maybe not'. Some targets are like that, unfortunately. But against those, is there any size of screening collection that would do the trick? There are definitely cases where DNA-encoded libraries or other similarly huge auxiliary collections (unusual peptides, for example) have delivered hits while the standard screening deck has not. That in itself is enough to make the we-need-more-compounds advocates feel as if they've won the argument. But no one is sure that if you screen, say, a totally disordered protein you should expect any real hits at all, no matter how many compounds you throw at it. Plotting hit rates of assays will give you a horribly long-tailed distribution, and no one knows how far it extends. Exploring it will take you up another unknown curve, where you keep spending more and more time and money with no way to predict a safe landfall.

So at one end of the range you can be nearly certain that you'll get chemical matter out of the screen; at the other, it may be that there simply is no such thing to get (I'm not sure I believe in that latter category myself, but neither do I have any proof that it doesn't exist). It's in that wide, middle ground that people can always complain about the

[†] This article was originally published in *Chemistry World*, first published online on 2 August 2018.

compounds they have... and dream
about the ones they don't.

Derek Lowe (@Dereklowe) is a medicinal chemist working on preclinical

drug discovery in the US. He blogs at *In the Pipeline*.