

## Databases aim to bridge the East-West divide of drug discovery

The pharmaceutical industry, like almost every other industry these days, is looking to China for growth. Over the past few years, numerous foreign drug companies, including GlaxoSmithKline, Novartis, Pfizer and Sanofi, have made significant inroads into the country by building manufacturing facilities and forging partnerships with domestic firms. But drug developers are also looking to the East for inspiration earlier in the pipeline. Several companies have active programs to screen traditional Chinese medicines (TCMs) for new drug leads, and now the introduction of large databases of molecular information about active TCM ingredients should make that process easier.

The latest database to hit the scene is Chem-TCM, which collates structural and pharmacological data from around 12,000 compounds found in more than 300 Chinese herbs. Although not a physical library of chemical entities, Chem-TCM—launched in October by researchers at King's College London in collaboration with the Shanghai Institute of Materia Medica—is part of a growing number of data repositories focused on plant-based traditional remedies that, organizers hope, will allow researchers to find new drug candidates through virtual screening and computational modeling.

Calvin Yu-Chian Chen, a bioinformatics researcher at China Medical University in Taiwan, was frustrated by the lack of two- and

three-dimensional structures of the chemical compounds found in traditional medicines. So he spent close to a decade developing his own catalog, dubbed TCM Database@Taiwan (*PLoS ONE* 6, e15939, 2011).

Chen's database is slightly larger than Chem-TCM, with records for over 20,000 compounds from 453 medicines. It is also freely available on the Web, whereas the King's College resource costs \$1,850 for academic investigators and about twice as much for commercial companies. "On the face of it, they're ahead of us," admits King's computational biologist David Barlow, one of the co-developers of Chem-TCM. But he notes that Chem-TCM might have a biological edge, because it lists pharmacological activity data along with structural information, whereas the Taiwanese database currently only shows molecular structures.

Even so, Chen has made the most of TCM Database@Taiwan. Since unveiling the resource at the beginning of the year, Chen and his colleagues have used it to identify new potential drugs targeted against influenza's hemagglutinin surface protein (*Mol. Biosyst.* 7, 3366–3374, 2011) and to find inhibitors of epidermal growth factor receptor, a major driver of cancer (*PLoS Comput. Biol.* 7, e1002189, 2011). "We always focus on the most important diseases," Chen says.

Although studies like these show the benefits of mining TCM-related data for drug compounds, some people see the

creation of separate TCM-specific databases as counterproductive. John Overington, who heads up ChEMBL, a million-compound database run by the European Bioinformatics Institute in Hinxton, UK, says that existing repositories derived from conventional chemical libraries are larger than TCM databases and, thus, better for finding the most potent compounds across all drug sources. "You get more power," he says.

Irene Kouskoumvekaki, a chemist who studies traditional Chinese remedies at the Technical University of Denmark's Center for Biological Sequence Analysis near Copenhagen, disagrees. She says that the standalone databases make it easier for researchers to link compounds back to the plants and traditional therapies from which they came. This information would probably be lost if the compounds were listed discretely in a larger database, she argues.

Regardless of where the data sit, Brian Shoichet, a computational chemist at the University of California–San Francisco, notes that TCM compounds must first become commercially available before screening their structures has any practical value. Most TCM ingredients are hard to come by, and, until the compounds listed can be bought to order, Shoichet sees TCM databases as simply a source of ideas, but not necessarily ones that are easy to translate into therapeutic applications.

*Katharine Sanderson*

## Biostatisticians call for more scientifically rigorous pilot studies

So-called 'pilot studies' often need more rigorous consideration before taking flight, according to a paper published in late October. The authors, a group of biostatisticians, call on researchers to exercise more care when designing and conducting these types of investigations (*Clin. Transl. Sci.* 4, 332–337, 2011).

When done right, pilot studies serve an important purpose; they usually involve small numbers of animals or participants and help researchers test the feasibility of methods that will be used in larger, more involved and expensive trials. Scientists can, for example, use pilot studies to test the tolerability of a given compound or the ease of trial subject recruitment. And often this information helps investigators net grant money to pursue larger projects.

But too often, the new paper argues, investigators see pilot studies as unimportant and fail to design them properly, and that attitude can harm investigators' careers and

stymie scientific progress. That's an especially pertinent problem for the research community, given that pilot studies seem to be on the rise. An informal search of the Medline database by Paul Nietert, a biostatistician at the Medical University of South Carolina in Charleston and a co-author on the new study, found the number of publications mentioning 'pilot study' in the subheading or as a key term has grown about threefold over the past two decades.

Lehana Thabane, director of the biostatistics unit at the Centre for Evaluation of Medicines in Hamilton, Ontario who published a similar paper on pilot studies last year (*BMC Med. Res. Methodol.* 10, 1, 2010), says many researchers fail to outline their criteria for success at the outset. Consequently, they have no measure by which to interpret their results.

Alternatively, some investigators try to do too much in a pilot study. Nietert has seen proposals in which the investigator outlines

plans to test a variety of hypotheses and conduct sophisticated statistical analyses on data from 15 people. "It's not going to be meaningful," he says. "They're probably underpowered to test all those hypotheses." Moreover, many pilot studies are never published. That means we can't learn from them, "and we keep repeating the same mistakes," Thabane says.

Notably, the term 'pilot study' is not often used by the US Food and Drug Administration, according to Lisa Kubaska, a spokesperson at the agency's Center for Drug Evaluation and Research in Silver Spring, Maryland. The agency "often sees proof-of-concept studies, but these are often combined with the phase 1 safety studies as a dual-purpose study." Nonetheless, background information can prove helpful, and "when such information does not already exist, well-designed pilot studies would be very useful."

*Cassandra Willyard*