

BY KATHARINE MILLER

Curating Drugs' Potential with SWEETLEAD

Pharmaceutical research is notoriously expensive. To find safe and effective drugs cost-effectively, some researchers seek new uses for medications that have already leaped the hurdles of the FDA approval process. One systematic approach to such drug-repurposing projects involves virtual screening of molecular structures to identify compounds likely to have a particular desired effect. But these efforts have uncovered a problem: "Different databases give different chemical structures for the same drug name," says **Paul Novick, PhD**, who recently completed his doctorate in Vijay Pande's lab at Stanford University.

Novick decided to address that problem by creating an algorithm that automatically evaluates the structures of existing medications in various public databases. The curated database he created, called SWEETLEAD, is described in the November 2013 issue of *PLoS One*.

Virtual screening is very sensitive to precise structural information, Novick says. "A good compound might rank really low and a crappy molecule with a wrong structure might score highly," he says. "Missing out on potential active compounds is a big concern."

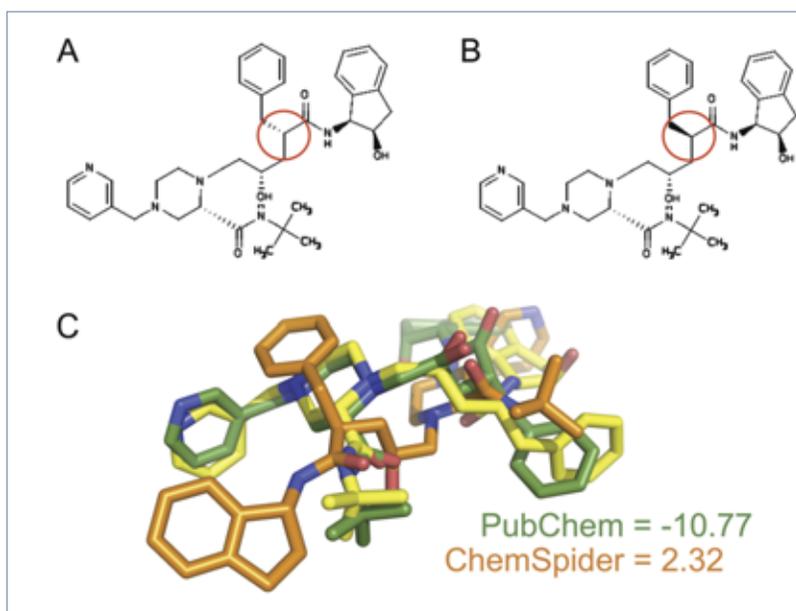
Unfortunately, the patent literature and regulatory documents that describe the molecular structure of existing medications are not currently available in downloadable form, Novick says. And reviewing that literature to re-enter all of the compounds manually would be both tedious and potentially ineffective. "It opens you up to the same kinds of errors that led to the original problem," Novick points out. "And as new drugs are approved, you want an automatic system for inclusion."

To create the SWEETLEAD database, Novick and his colleagues started by querying multiple databases (PubChem, ChemSpider, DrugBank, and others) for the chemical IDs that match a particular drug or herbal isolate's name. The algorithm then compares the structures for those IDs to see if there is a majority or consensus structure. If yes, then SWEETLEAD tags the name to that structure. For drugs with no majority or consensus structure, Novick manually reviewed the patent literature and then tagged the accurate structure.

Novick concedes that there is no *de facto* reason to trust majority structures except that they are well used by researchers who are highly motivated to correct errors. But as a final check on SWEETLEAD's accuracy, Novick compared the structures to a several other databases. "Where there were discrepancies, our structures were accurate more often than theirs," he says.

The SWEETLEAD database includes 3,600 molecules, including 2,000 approved drugs, many recreational drugs, and numerous chemical isolates from traditional and herbal medicines. "These represent a good starting point for further study by anyone doing repurposing projects," Novick says.

In addition, Novick says, SWEETLEAD can be used to explore commonalities among approved drugs. For example, the database can be used to challenge the rules-of-thumb (such as Lapinski's rule of five) that many pharmaceutical researchers use to define whether a molecule is drug-like or not. "Researchers frequently ignore compounds that violate these rules, missing out on potentially active compounds," Novick says.



The structure of indinavir, a protease inhibitor approved for treatment of HIV and AIDS, exhibits different stereochemistry (red circles) in PubChem (A) compared to ChemSpider (B). The PubChem structure was correct and received a high score for its potential to inhibit HIV protease (C) while the incorrect structure from ChemSpider received a low score.

Novick has already used SWEETLEAD to identify several compounds that are a few steps away from clinical trials, including one for treating Chagas disease and another for Dengue fever. He's hopeful they will be effective at the same dose for which they are already approved, which would allow them to skip Phase I clinical trials.

But even if these efforts don't pan out, Novick says, "From a drug discovery perspective, any compound from our database identified as a drug candidate would definitely be a sweet lead." □

Simbios (<http://simbios.stanford.edu>) is the National Center for Physics-Based Simulation of Biological Structures at Stanford.

simbios

DETAILS

SWEETLEAD is publicly available at simtk.org/home/sweetlead.