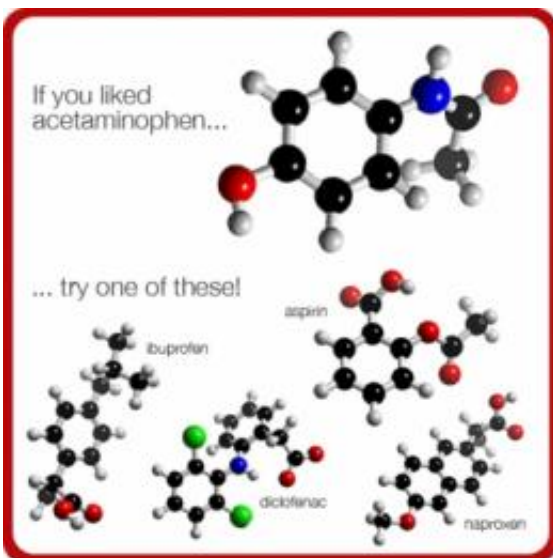# Drug discovery, Netflix style?

April 14 2010, by Larry Hardesty

Graphic: Christine Daniloff

(PhysOrg.com) -- In the last 10 years, the growth of the Internet has made ranking algorithms one of the hottest topics in computer science. The most famous ranking algorithm is Google's, which determines the order of search results, but close behind are the Netflix and Amazon algorithms that make recommendations on the basis of customers' prior decisions. Now researchers at MIT and Harvard Medical School have shown that ranking algorithms could find an important application in a somewhat surprising field: drug development.

Drug development typically begins with the identification of a "target" — a molecule involved in the biological processes underlying some

disease. The next step is to try to find chemicals that either promote or suppress the molecule's production. Scientists have assembled huge libraries — both virtual and physical — of chemical compounds that might be active against biological targets, and drug developers who have identified a target usually select a group of candidate drugs from those libraries.

But the majority of drug candidates fail — they prove to be either toxic or ineffective — in clinical trials, sometimes after hundreds of millions of dollars have been spent on them. (For every new drug that gets approved by the U.S. Food and Drug Administration, pharmaceutical companies have spent about $1 billion on research and development.) So selecting a good group of candidates at the outset is critical.

Drug companies have been using artificial-intelligence algorithms to help select drug candidates since the late 1990s. But in a paper appearing in the next issue of the American Chemical Society's *Journal of Chemical Information and Modeling*, Shivani Agarwal, a postdoctoral associate in the Computer Science and Artificial Intelligence Laboratory, Deepak Dugar, a graduate student in chemical engineering, and the Harvard Medical School's Shiladitya Sengupta showed that even a rudimentary ranking algorithm can predict drugs' success more reliably than the algorithms currently in use.

At a general level, the new algorithm and its predecessors work in the same way. First, they're fed data about successful and unsuccessful drug candidates. Then they try out a large variety of mathematical functions, each of which produces a numerical score for each drug candidate. Finally, they select the function whose scores most accurately predict the candidates' actual success and failure.

The difference lies in how the algorithms measure accuracy of prediction. When older algorithms evaluate functions, they look at each

score separately and ask whether it reflects the drug candidate's success or failure. The MIT researchers' algorithm, however, looks at scores in pairs, and asks whether the function got their order right.

"The criterion we're giving it is almost the simplest ranking criterion you could construct," Agarwal says. Nonetheless, in experiments involving data on existing drugs, it consistently predicted the drugs' success more reliably than the algorithms now in use. The improvements were relatively modest, but to Agarwal, they're an indication that recent research on more sophisticated ranking algorithms holds real promise for drug discovery.

"There's a really very systematic improvement over previous methods, and that's quite striking," says Peter Bartlett, a professor of computer science and engineering at the University of California, Berkeley. "This is a very nice empirical demonstration that these methods are more effective than the standard methods."

Anton Hopfinger, a professor at the University of New Mexico College of Pharmacy, cautions that when computer systems rank drug candidates, "the key component is not too surprisingly the properties of the drug or molecule you use to train the system." That is, the success of the system depends crucially on the mathematical descriptions of the drug candidates. Even the ideal algorithm is helpless if it's acting on data uncorrelated with a molecule's biological activity.

But Agarwal is a computer scientist, not a biologist. So while the biologists continue to refine their descriptions of the chemical properties of biological molecules, Agarwal continues to refine her algorithms for ranking drug candidates. At the moment, she's investigating algorithms that maximize the accuracy of the rankings at the top of a list, even at the expense of lower rankings, since drug developers are generally interested in only a handful of the most promising drug candidates.

More information: Journal paper:

Provided by Massachusetts Institute of Technology