In Brief

Across the globe, Molplex is helping scientists reduce the time and cost of screening large data sets of chemical compounds to identify potential drugs. Researchers are using automated tools in the cloud to increase the rate of success in drug discovery as they work to treat tropical diseases such as malaria, tuberculosis, and dengue fever.

David Leahy
Chief Executive Officer
Molplex

Vladimir J. Sykora
Chief Operating Officer
Molplex

Paul Watson
Professor of Computing Science
Newcastle University

Fabrizio Gagliardi
Cloud Engagements in EU
Microsoft Research Connections

Websites:

research.microsoft.com/cloudresearch

windowsazure.com

molplex.com

ncl.ac.uk

Aiming to Deliver New Drugs Faster at Less Cost in the Cloud

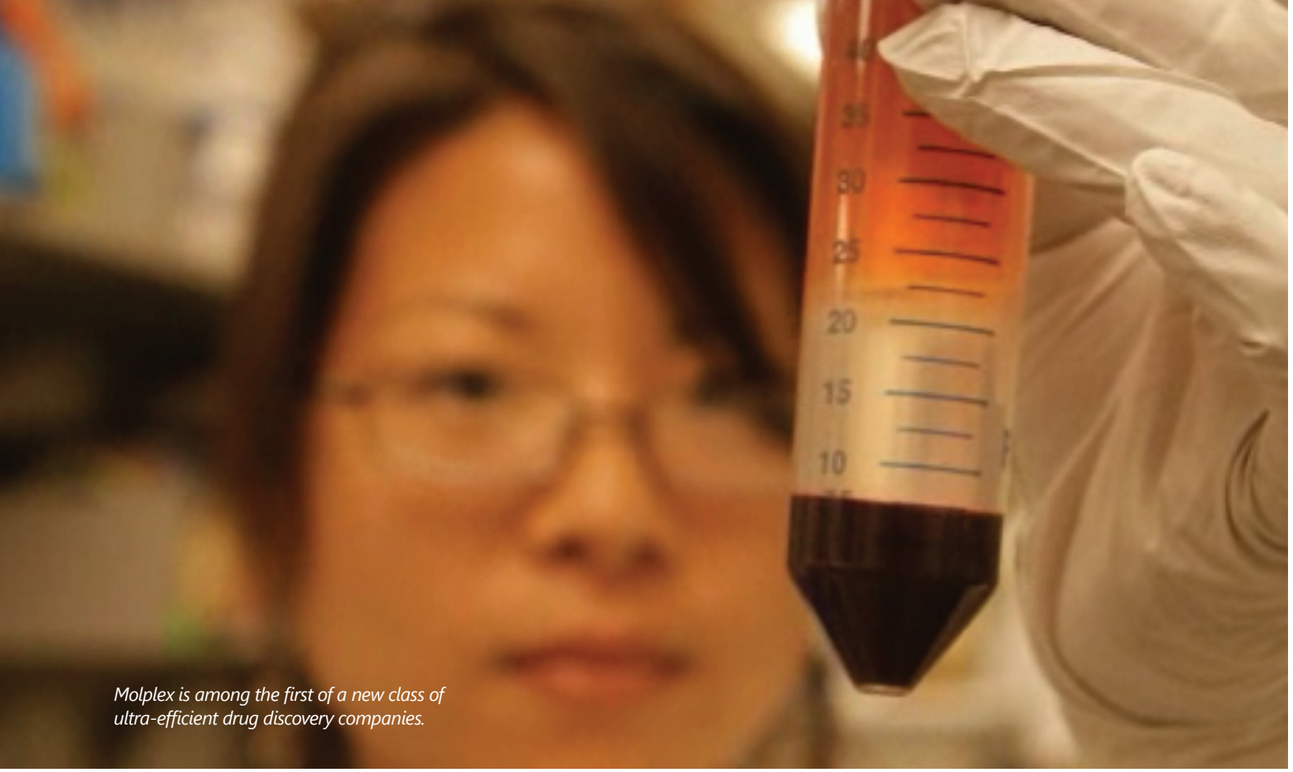
Scientists worldwide are attempting to discover new medicines for more diseases than ever before. Modern drug discovery involves identifying likely candidates and screening them to find effective biologically-active compounds. To find promising "hits" quickly from millions of candidates presents a technical challenge for researchers. Microsoft Research has a technological solution to the issue: Molplex Clouds Against Disease.

Toxicity prediction remains one of the great challenges of drug discovery. Despite decades of unprecedented funding, scientists still cannot predict all possible toxic side effects for a given drug. Traditional statistical models based on empirical data, while useful in theory, have one key shortcoming. Unless researchers have access to either a state-of-the-art corporate data center or one of the world's few supercomputers, there's just too much data to analyze efficiently. The identification of compounds that will cause a desired biological effect requires a huge investment in technical infrastructure.

Researchers from Molplex, a small drug discovery company; Newcastle University; and Microsoft Research Connections are working together to help scientists across the globe deliver new medicines more quickly and at lower cost. This collaborative partnership has helped Molplex develop Clouds Against Disease, an offering of high-quality drug discovery services based on a new molecular discovery platform that draws its power from cloud computing.

RETHINKING DRUG DISCOVERY

David Leahy, co-founder and chief executive officer of Molplex, envisions a way to help pharmaceutical researchers anywhere in the world form effective drug discovery teams without large investments in technology or fixed running costs. "It takes massive computing resources to search through chemical and biological databases looking for new drug candidates. Our Clouds Against Disease solution dramatically reduces the time



Molplex is among the first of a new class of ultra-efficient drug discovery companies.

and cost of doing that by providing computation and chemical analysis services on demand," Leahy says.

Molplex is among the first of a new class of ultra-efficient drug discovery companies. Clouds Against Disease frees smaller companies from the up-front capital costs for what is typically an overprovision of computing resources. Public and private entities alike can create high-value strategic drug portfolios without the financial strain of investing in costly physical

"Using our platform to identify antibacterial leads that would be effective against methicillin-resistant staphylococcus aureus (MRSA), we went from an idea for intervening in bacterial growth through to a set of compounds that are actually killing the bacteria in just five weeks and at a cost of less than $2,000."

David Leahy
Chief Executive Officer
Molplex

technology infrastructure. Leahy explains, "Imagine being able to lower the cost of drug discovery from $30 million to $1 million. It changes the risk profile of early drug discovery."

Molplex regards drug discovery as a big data and search optimization problem. Clouds Against Disease uses its computational molecular discovery platform to automate decision making that is traditionally the scientists' task.

"Instead of having teams of scientists scanning chemical information, our software searches for structures that have multiple properties matching the search criteria," explains Leahy. "When we integrate that with highly automated chemical synthesis and screening, it becomes a much more efficient and productive way of doing drug discovery."

DATA MANIPULATION ON A LARGER SCALE

In a recent pre-clinical study, the company applied its computational platform to more than 10,000 chemical structure and biological activity data sets. This action generated 750,000 predictive relationships between chemical structure and biological effect. After generating numerous possible outcomes, Molplex then used the same validation criteria that scientists would use to narrow down the 750,000 relationships to just 23,000 models covering 1,000 biological and physico-chemical properties, a relatively small data set that humans could then manage. "It would have taken hundreds of scientists several years to do this the conventional way," Leahy says.

Once the models were built, the Molplex system used them to sift through a massive database of new chemicals to seek compounds that are active against bacterial infection. By using this combination of data manipulation features, Molplex has generated new leads against tropical diseases such as malaria, tuberculosis, and dengue fever. Leahy notes, "With computing power from Windows Azure, our drug discovery process took only three weeks on the very first pass, and has become even quicker more recently."

The average costs of drug discovery are so high not because of the project costs, but because most projects fail to deliver anything. "Our process features extreme front loading that identifies viable drug candidates early, while researchers traditionally do a lot of experimental work and then find out late in the day that the compound is toxic," Leahy notes. The Molplex method allows researchers who are screening compounds to address practical questions, such as: Will the compound be toxic? Will it pass safely through the human intestine? Will it stay in the body long enough?

Windows Azure, Microsoft's cloud platform, was critical to the success of Clouds Against Disease. Molplex can access 100 or more Windows Azure nodes—in effect, virtual servers—to process data rapidly. The physical-world alternative would be to source, purchase, provision, and then manage 100 physical servers, which represents a significant investment in up-front costs. Before they could begin drug research, scientists taking this traditional approach would have to raise millions of dollars, but Windows Azure helps eliminate start-up costs by allowing new companies to pay for only what they use in computing resources.

A new technology that contributes to the efficient use of cloud computing, Generic Worker is a worker-role implementation for Windows Azure that enables multiple drug discovery teams to use the Molplex systems concurrently. Research projects are distributed and scaled across a set of Windows Azure nodes. But researchers do not need to understand the technical details of the infrastructure or learn new applications for their daily work. They simply use Generic Worker to submit their jobs to Windows Azure and receive the processed results.

"When researchers had to depend on older servers in a data center, we estimated that it could take up to five years to find a bacteria-killing compound. Now that researchers can use up to 200 nodes on Windows Azure with 90 percent efficiency, that same process can be accomplished in about 10 hours. Imagine what that means for the future of drug discovery."

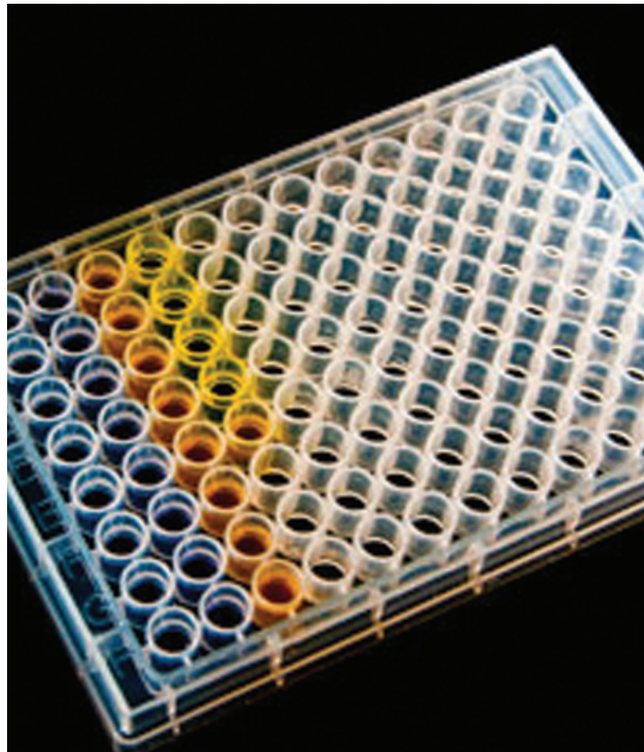
Paul Watson
Professor of Computing Science
Newcastle University

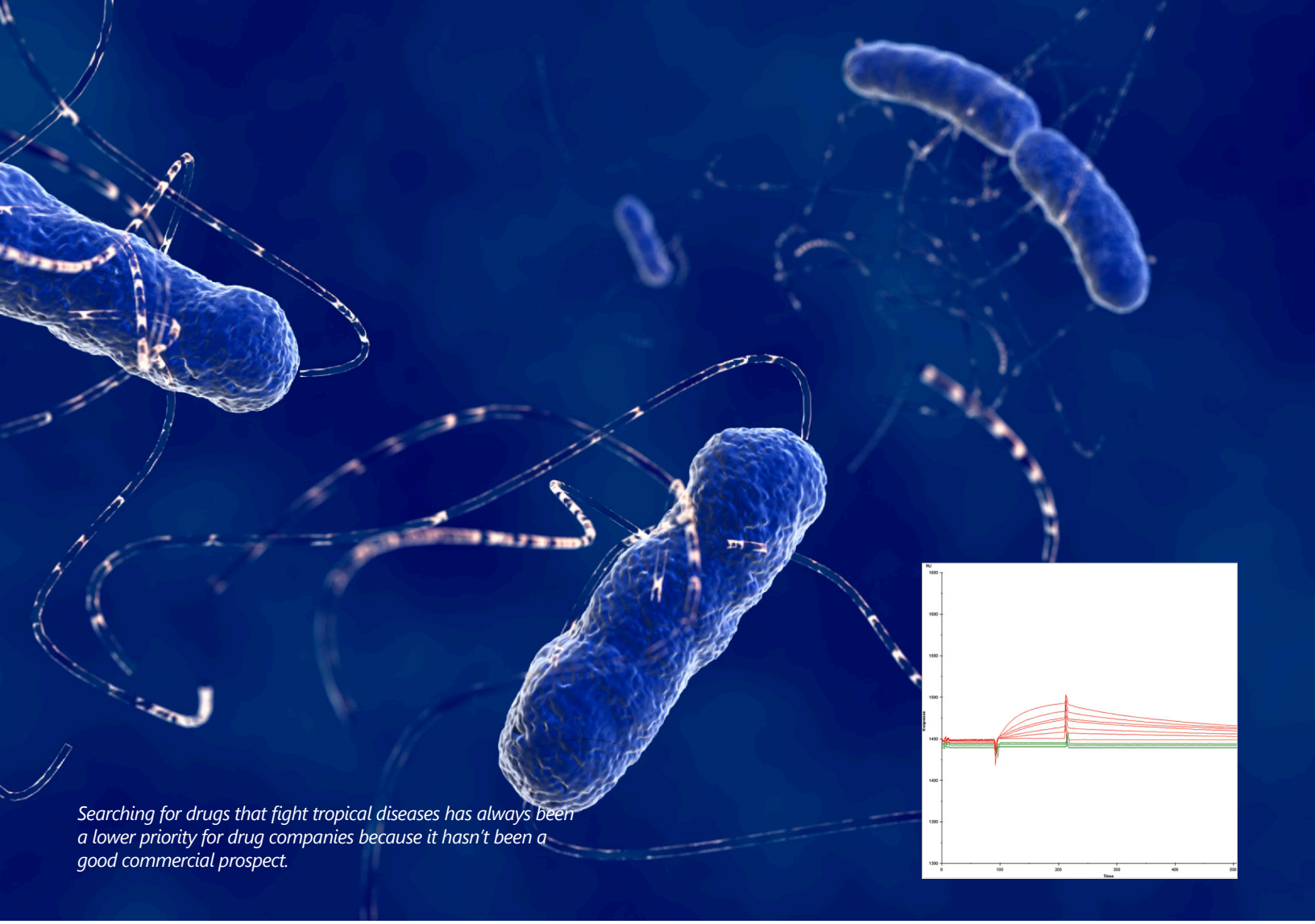
Vladimir J. Sykora, co-founder and chief operating officer for Molplex, is the primary end user of Clouds Against Disease. Sykora explains that the Molplex computational platform runs algorithms his company developed to calculate the numerical properties of molecules rapidly. Consequently, Molplex has been able to produce drug discovery results on a much larger scale than what was previously feasible. "We would not have been able to predict so many compounds without the cloud computing resources enabled by Windows Azure," asserts Sykora. "The speed and high level of detail provided by Windows Azure allow us to explore far beyond what would have been possible with traditional hardware resources."

DEVELOPMENT OF CLOUDS AGAINST DISEASE

Clouds Against Disease originated from a highly productive collaboration between Newcastle University and Microsoft Research Connections. The impetus for the partnership came from Paul Watson, professor of Computing Science at Newcastle

Once the models were built, the Molplex system used them to sift through a massive database of new chemicals to look for compounds that are active against bacterial infection.





Searching for drugs that fight tropical diseases has always been a lower priority for drug companies because it hasn't been a good commercial prospect.

University, who is acknowledged as a world expert in cloud computing. "Watson recognized the potential of cloud computing to address the big data problem of drug discovery," says Leahy. "Because of his excellent relationship with Microsoft, he saw Windows Azure as a solution to that."

Clouds Against Disease came about through an academic collaboration that began when Leahy and Sykora worked with Watson as part of the university's computing research team. Soon after the 2009 announcement of Microsoft SQL Azure Database, the three began looking for ways that cloud computing could help science. Their paths then diverged for a time.

Watson and his other university colleagues decided to take part in the VENUS-C cloud infrastructure project. Funded by the European Union, the VENUS-C program was created to provide computing resources, such as Windows Azure, to scientific users and small start-ups whose businesses required massive, on-demand computing resources. Meanwhile, seeing the commercial potential of cloud computing for drug discovery, Leahy and Sykora left the university to start Molplex.

The three were united again when the VENUS-C consortium decided to nominate Molplex for participation in VENUS-C as a pilot project. "Molplex and Clouds Against Disease were a great fit for the VENUS-C pilot, because their business plan called for using virtual servers in the cloud to find chemical compounds to treat diseases," explains Watson.

Leahy notes, "VENUS-C served as an incubator that helped us quickly move from a theoretical concept to a commercially viable endeavor. Access we gained to the Windows Azure platform through VENUS-C was essential to our ability to demonstrate the feasibility of our business model."

FIGHTING TROPICAL DISEASES

Molplex is embarking on a new collaboration with the Malaysian government to search for drugs that fight tropical diseases. This search has always been a lower priority for drug companies because the market is smaller, making it a less desirable commercial prospect. The traditional drug discovery program is geared to $1 billion a year blockbuster drugs; however, there are fewer opportunities today for drugs with that level of commercial potential.

Increasingly, scientists are researching tropical diseases that affect smaller populations; radically reducing the cost of drug discovery makes it feasible for scientists to tackle them. "Unlocking drug discovery technology from a physical location with the cloud has tremendous potential to help researchers work on curing these diseases faster and at less cost," asserts Leahy, "wherever they are in the world."