Where should the drug industry go to find new ideas? In the first of two features, **Alison Abbott** asks if the future lies in systems biology — a field that attempts to piece together 'everything'. In the second, **David Cyranoski** looks at drug companies' attraction to China.

A FIENDISH PUZZLE

If it is hard to make something from nothing, then it can be just as hard to make something from everything. But that, in essence, is what many pharmaceutical companies are trying to do as they seek new drug targets by integrating the massive sprawl of biological information now available.

The puzzle is like a jigsaw with an unknown number of pieces and, as yet, no edges. So when Cristiano Migliorini of the drug giant Roche saw a promising drug target for type 2 diabetes emerge from hundreds of millions of data points, he felt a celebration was in order. He picked up the phone and invited his 20 or so collaborators from research institutions all over Switzerland to a party.

The celebration — a good dinner, drinks and some lively pub games — might have been considered premature. It is a long and tortuous route from an interesting protein to a new and effective drug. But since the party, this protein has been put through its paces at Roche's drug-development labs in Basel, and candidate drugs that bind to it are already lined up for testing. The target could still fail, but in Migliorini's mind, the results have already shown that the mega-data-crunching approach of 'systems biology' — new, modish and increasingly adopted by pharmaceutical firms — can piece together a meaningful picture from this colossal biological puzzle. "And that," says Migliorini, "is worth at least a good meal and a game of pool".

The crisis-ridden pharmaceutical industry desperately wants something to celebrate. The glory days of the blockbuster drug seem to be over. When genomics matured at the turn of the century, much of the industry was convinced that individual genes would emerge as the new drug targets. But that reductionist bubble soon burst: the more that geneticists and molecular biologists have discovered, the more complicated most diseases have become. As individual genes have fallen out of favour, 'systems' — multitudes of genes, proteins and other molecules interacting in an almost infinite number of ways — have come into vogue. Systems biology is an attempt to make sense of all these data.

Some researchers and analysts were cynical when systems biology was hyped as the saviour of the failing research and development pipelines. "Rightly so, perhaps," says Giulio Superti-Furga, who studies complex protein interactions and is head of the Research Center for Molecular Medicine in Vienna. "A few years ago, systems-biology proponents were writing reviews that triggered unreasonable expectations for advances in medicine." Many companies have therefore been cautious about making a big investment. Like Roche, most are testing the water with defined projects, usually in collaboration with top academics or small computational companies.

Although systems biology is far from proven as a drug-discovery tool, it has many advocates and its wider adoption seems to be inevitable. If leading biologists are concluding that disease can only be understood by tackling the entire, messed-up system, pharmaceutical companies arguably have little choice but to keep up. "Industry really needs to dive in," Superti-Furga says.

Never-ending networks

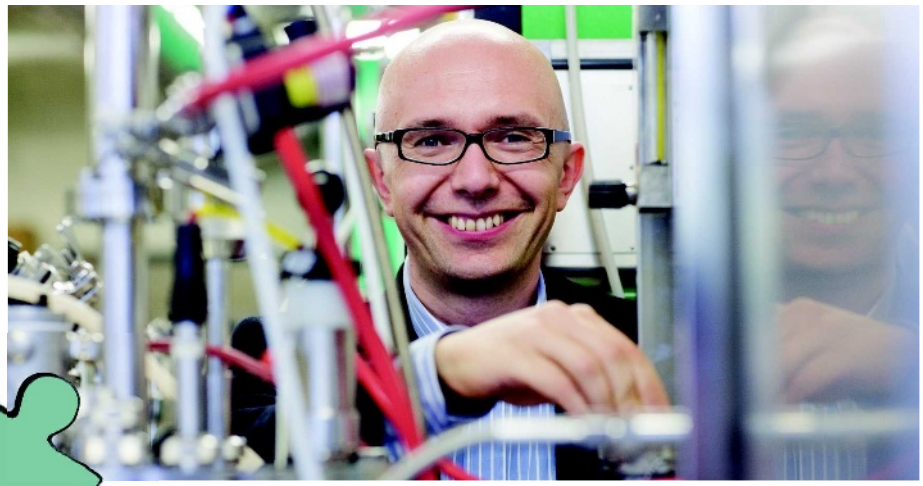
At its heart, systems biology is about gathering unprecedented amounts of data from cells, then making sense of it through mathematical models. At its most sophisticated, it might involve the high-throughput collection of molecular data, such as DNA sequences, RNA molecules, proteins and metabolites as well as more descriptive data such as clinical diagnoses and drug responses. These data are then assimilated into computational models of cellular processes which, as molecules change location and function every microsecond, must also accommodate the dimensions of time and space.

It is not enough to mimic what's known about the cell — such models must also predict what is unknown so that scientists can test their hypotheses. A model of a liver cell, for example, might reveal which of 100 interesting drug candidates will prove beneficial and which will be toxic. Knowing when to pull a doomed drug from the pipeline could spare a company a lot of expensive work. "The biggest value of systems biology immediately is as an aid to decision-making," says David de Graaf, who has worked in systems biology programmes at Pfizer, AstraZeneca and now at Boehringer Ingelheim. "And a drug company is a decision-making machine."

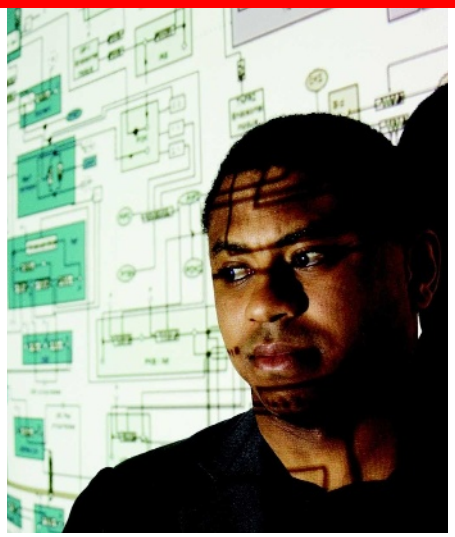
Although systems biologists want it all, for now they cannot get it. "We'll not have everything a hundred years from now," says Colin Hill, chief executive of the biotech company Gene Network Sciences in Cambridge, Massachusetts, which uses supercomputers to model complex biology for drug companies. "Even after two decades of gathering and

ILLUSTRATION: J. HEIN VAN DIERENDONCK

C. ROTH



Cristiano Migliorini has seen systems biology deliver promising drug targets.



J. RADIN

Colin Hill uses his background in theoretical physics to model complex biological systems.

archiving molecular data, barely 5% of the total circuitry inside the cell is known." So systems biologists must choose between two approaches when developing their models. In one, they scale back their ambition, paring down the number of biochemical pathways being modelled to just a few well-known ones. In the second, they bluff the bits they do not know, trying to model 'everything', and using sophisticated computational tricks to compensate for absent data.

At the end of the day, the choice of approach depends on what question researchers want to ask of their models. It's a bit like the conundrum in Douglas Adams's *Hitchhiker's Guide to the Galaxy*, where the computer Deep Thought spends 7.5 million years computing an answer to the most horribly imprecise "ultimate question of life, the Universe and everything" — only to come up with the answer "42". If the question is narrow — 'which surface proteins are affected when insulin activates receptors on the pancreas cell?' — then it may be sufficient to model only those networks known to be involved in insulin signalling. If the question is broad — 'what would happen to an animal fed drug X?' — then the 'everything' approach is more appropriate to see what falls out, be it stronger muscles or liver failure.

The meaning of life

Drug companies are divided on which of the two approaches is better, and Migliorini is clear about which he prefers. "We take modelling in small doses — a defined number of well-understood pathways," he says. "We have a pragmatic approach — we can't say '42' to management." Cell biologist Willy Krek, an academic partner of Miglorini's at one of Switzerland's foremost universities, the ETH in Zurich, adds: "It is strategically better to get a rich data set on a limited network than to take measurements for the rest of your life, always saying they are still not enough."

> "We'll not have everything a hundred years from now."
> — Colin Hill

Even the scaled-back approach to systems biology takes massive effort. Migliorini's collaboration began in 2006 as part of Switzerland's big systems-biology research programme now known as SystemsX.ch. Migliorini calls it "an industry–academic collaboration beyond cliché", referring to what is in fact the largest such partnership between pharmaceutical companies and academia in biology that Switzerland has ever seen. It involves several top Swiss research groups, plus academic and industrial partners elsewhere, at a cost of 9 million Swiss francs (US$8 million) over three years as well as undisclosed internal spending.

In searching for its diabetes drug, the consortium has built a model focused on the 500 or so proteins — called the 'surface proteome' — on the coat of pancreatic β cells, the cells that secrete insulin when blood glucose concentrations rise. Migliorini and his coworkers are working on the concept that some people who overeat develop type 2 diabetes because their β cells cannot keep up with the body's demands. The team is seeking to identify surface proteins that stop responding to environmental signals such as glucose metabolites, as these may be responsible for transforming the overworked cells into diseased cells.

The scientists use almost every high-tech method imaginable to gather data on proteomes and the cellular networks they link into. Consortium member Rudi Aebersold from the ETH developed one such method using mass spectrometry that separates out glycoproteins — the proteins involved in key processes such as protein folding or cell–cell signalling — from the background of other proteins[1]. Importantly, the technique quantifies as well as identifies proteins, providing a molecular fingerprint for a cell in a particular stage of health or disease. Quantification is essential to systems biology, whereas in the past, molecular biologists were happy just to see whether or not a protein was present.

To flesh out the proteomic information, the researchers also generate reams of other data, either experimentally or by pulling them out from public and commercial databases, to feed into their computational model. They then work out which surface proteins converge on the same intracellular pathways. At the end, they have a massive, moving, knitting-ball of a network in which every protein is linked to every other protein it interacts with. The hope is that this can point to the 'nodes' — the key points in a network that have disproportionate influence.

Then researchers can ask questions of the model. For example, if the team suspects that the disease is being caused by a particular aberrant glycoprotein, they strip that protein from the model. If correct, the model should then mimic the knock-on effects in the circuitry that are known to occur. That glycoprotein then becomes a potential drug target or, perhaps, a biological marker of disease progression.

Quick results

At the start, not everyone at Roche expected the systems approach to succeed within the three years the collaboration was initially given. In fact, Migliorini's first hit, last autumn, came within a year of his project starting. "It was a great feeling to see our engineering-style approach — multiple inputs, each comprising amounts of data that make you sweat — spit out a clear single target," he says. Here, luck intervened: the protein in question (its identity is still guarded) had been investigated in relation to an entirely unrelated disease, so the tools and compounds for studying it were already available. "The whole process went lightning-fast," says René Imhof, director of research for Roche in Basel. Roche's academic partners are also happy, because the collaboration was formed at exactly the time they were starting to feel the limitations of the traditional genes-will-explain-disease reductionism.

Across the Atlantic, in the Boston area of Massachusetts, Peter Sorger and Douglas Lauffenburger have been feeling similar constraints. Sorger says that he used to regularly raise hackles with his outspoken criticism of traditional genomics. As the region is home to the Whitehead Institute for Biomedical Research, one of the most important sequencing hubs of the Human Genome Project, those hackles were being raised on some particularly high-level necks.

Sorger and Lauffenburger now operate complementary systems-biology labs — Sorger at Harvard Medical School in Boston and Lauffenburger at the Massachusetts Institute of Technology in Cambridge — and have been asked to collaborate with numerous pharmaceutical companies, usually in the area of cancer. Both have worked with AstraZeneca, for instance, to understand the action of the company's anticancer drug gefitinib (Iressa). Gefitinib was a much-heralded drug designed to hit the ERBB1 receptor when it is mutated, as it frequently is in various cancers. But few patients responded to the drug, and genetic sequencing showed that mutations in the

receptor could not predict who these patients would be. "By focusing down on just the gene, the genomic approach had taken the target out of its physiological context," says Sorger, "and the physiology needed to be put back with systems biology."

Lauffenburger attempted to do this by looking at the whole life cycle of the errant receptor and how it connected with biochemical pathways inside the cell. Like many membrane receptors, ERBB1 switches off its signalling capacity by removing itself from the surface of the cell, and Lauffenburger theorized that sensitivity to gefitinib could depend on how well this process works. His team collected data on genes, transcription, protein phosphorylation and the internalization of the receptor from cells that were either sensitive or insensitive to gefitinib and built a model to mimic the circuitry associated with the receptor.

"Physiology needed to be put back with systems biology." — Peter Sorger

The model suggested that mutations in ERBB1 were almost irrelevant: gefitinib would work only in cells that had inefficient internalization of ERBB1 (ref. 2). So far this has not generated new biomarkers for the company's clinical trials, but it has shown that complicated systems biology may be needed to select patients for drug treatments.

Like Migliorini and many others in systems biology, Lauffenburger has an engineering background. Others, such as Hill, have swept in from theoretical physics. Hill turned to biology after a stint at the interdisciplinary Santa Fe Institute in New Mexico, where he realized that physicists like himself had become comfortable contemplating the vastness of the Universe or the minuscule nature of the subatomic world, but were distinctly uncomfortable with a mid-sized affair such as life. "Life was always too complicated a subject for us," he recalls.

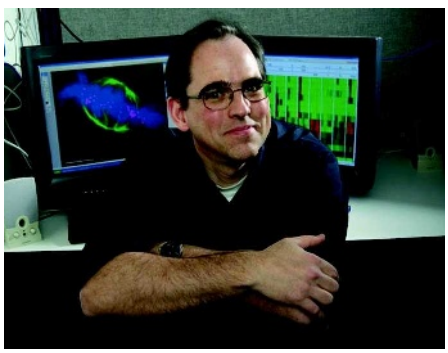
Hill's type have brought with them the sort of Bayesian, or probabilistic, analyses that have been applied in other areas that have huge and incomplete data sets, such as astronomy. They use these techniques to tackle the second approach to systems biology, relying on computer models to fill in the gaps between relatively sparse data points. Researchers then put real data into the model and modify it to fit the data better, repeating these adjustments until all possible permutations have been explored.

This type of modelling requires a scale of computing effort analogous to that required to predict weather and understand global warming. The number of ways in which one gene or protein can link to another in the tangled web of intracellular pathways and networks is astronomical. "If you take 22 variables, such as gene, transcript or clinical measurements, they may each affect an average of maybe 250 points in a network — that means there could be 250^{22} possible combinations," says Hill. "We can only do this with supercomputing." Hill's company is now cranking out potential drug targets, which it patents and then peddles to pharmaceutical companies.

Industry giant Merck is known for its bold move to incorporate the 'everything' approach into systems biology into its drug-development programmes. In 2001, it acquired the Seattle-based genomics company Rosetta Inpharmatics to develop systems genomics in house. "The scale of the biological information we have today is unprecedented," says Eric Schadt, Rosetta's science director and a former physicist. "Fifty thousand or so transcripts, millions of mutations, hundreds of clinical endpoints — a high-performing computing environment is essential and we have one of the world's biggest."

Growth factors

Earlier this year, the group used the model to show that its systems approach had identified key disease-causing gene networks in metabolic diseases[3]. The paper was published back-to-back with another that had used a systems approach to identify three obesity-related genes that had not sprung out of genomics data alone[4]. "The surprise here was to find that not just a handful of genes were causally implicated in the disease, but hundreds — whole pathways were involved," says Schadt. "So the challenge is to hit the nodes in networks where implicated pathways overlap." The company says that up to 30% of targets for various diseases in its early pipeline have been identified through Schadt's techniques, and that a few are in early-phase clinical trials.



MIT

Systems biologist Peter Sorger has criticized genomics techniques.

But some people in industrial and academic circles remain hesitant about systems biology. No one can be sure that it will really increase the number of targets or biomarkers that make it through clinical testing. Still, it is a gamble that almost all companies seem willing to make, even if investment levels are sometimes small. In a report published earlier this year, industry analysts at PriceWaterhouseCoopers argue that the pharmaceutical industry needs to rely much more on systems biology if it is to survive the failing-pipeline crisis and predicts that the approach will have become more prevalent by 2020.

Stephen Friend, Merck's vice-president for oncology, thinks that any hesitancy will be overcome when the modelling becomes so predictive that the toxicity and efficacy of a potential drug can be forecast very accurately even before an experimental animal is brought out if its cage. "The next three to five years will provide a couple such landmark predictions and wake everyone up," he says.

The challenge for systems biologists is that the system is being stretched at all ends of the imagination. There will inevitably be new categories of biological information to collect and feed into models. Measurements will be made on finer and finer scales so that individual molecules can be identified and counted in single cells with greater speed and efficiency. This information will be integrated into grander models, although hopefully not ones that take 7.5 million years to compute answers. At a meeting in Japan in February, scientists signed up to a declaration calling for a grand challenge to create a virtual representation of the physiology of the entire human within three decades. To get that far that fast will require formidable advances in technology development, and even greater computational power and confidence.

But for the time being, most systems biologists are happy to have their questions answered, or at least considered, by drawing boundaries around 'everything' such that it is contained within the walls of a virtual cell. "In fact, I don't think we have a choice," Migliorini says. Life, the Universe and almost everything is about the most that biologists can handle for now. ■

Alison Abbott is *Nature*'s senior European correspondent.

1. Zhang, H., Li, X. J., Martin, D. B. & Aebersold, R. *Nature Biotechnol.* **21,** 660–666 (2003).
2. Hendriks, B. S. *et al. IEE Proc. Syst. Biol.* **153,** 22–33 (2006).
3. Chen, Y. *et al. Nature* **452,** 429–435 (2008).
4. Emilsson, V. *et al. Nature* **452,** 423–428 (2008).