

# Betting on tomorrow's chips

At the proteomics frontier, dozens of companies are trying to develop the protein equivalent of DNA microarrays. But designing these chips poses much tougher technical challenges, says Alison Abbott.

**T**he genomics revolution has already brought us DNA chips — microarrays that consist of short DNA sequences immobilized on a surface. By determining which spots bind to messenger RNA (mRNA) extracted from a biological sample, geneticists can obtain an instant snapshot of the activity of thousands of genes at a time.

DNA microarrays are transforming studies of gene expression. But some scientists are already dreaming of the chips of the future, which they argue will carry tens of thousands of protein 'capture' molecules, each geared to identify and bind to one particular protein. With an appropriate detection system, such chips would be even more valuable than their DNA counterparts. Proteins, after all, are the business end of gene expression and the usual target of drugs. A chip that could simultaneously analyse the production of tens of thousands of them would be a boon, both to those engaged in fundamental research and to the drug industry.

Dozens of companies are already working on technologies to make the chips, and the hype is feverish. But sober analysis indicates that the technical hurdles to be overcome are so great that over the next couple of years we can only expect chips with fewer than 100 capture molecules; the handful of chips marketed so far carry fewer than 10.

Such low-density chips are fine for certain

applications, such as simple medical diagnostics. But for large-scale proteomics projects that aim to determine how complex patterns of protein production vary with disease, they are inadequate. And although some companies claim to be making good progress towards chips with tens of thousands of capture molecules, many experts are sceptical.

"The hype from extrapolation from DNA arrays is very harmful — the expectations are too high," says Richard Mason, a business-alliances analyst with the British company Cambridge Antibody Technology (CAT). "Proteins are much more complicated, and development costs will be orders of magnitude greater," he argues.

## The protein poser

For DNA chips, designing capture molecules and developing read-out systems were relatively easy. Strands of DNA bind tightly and specifically to mRNAs with a 'complementary' sequence. And if the mRNAs in a sample have all been tagged with a fluorescent dye, determining where on the array they have bound is also simple.

Proteins pose a much tougher challenge. Specific capture molecules must be designed for all possible proteins encoded by the genome — and also for the modified forms produced by processes such as phosphorylation or the addition of sugar groups.

Whereas the binding between DNA and mRNA is highly specific, finding a capture molecule that will bind with high affinity to one protein alone is extremely difficult.

The classical capture molecules for proteins are antibodies, which are themselves proteins, and most of the companies in the protein-chip business are working with them. "So far, antibodies are the only capture molecules that have been demonstrated to work at high specificity and sensitivity," says Leigh Anderson, chief scientific officer of Large Scale Biology, which has its proteomics division in Germantown, Maryland.

Nowadays, large libraries of antibodies can be produced using a procedure called phage display. CAT, which has been manufacturing antibodies since 1990, has cloned billions of distinct antibody genes from white blood cells of healthy individuals, and has inserted them into viruses called phages that infect *Escherichia coli* bacteria<sup>1</sup>.

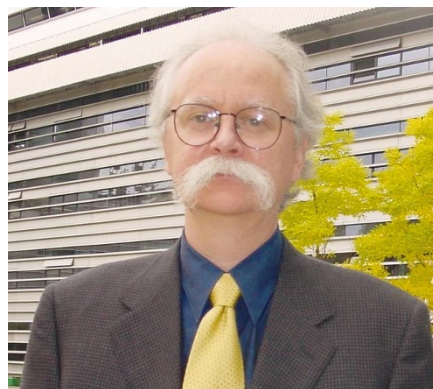
The phages reproduce in cultures of the bacteria. Infected cells eventually rupture, releasing phages into the growth medium. The remaining bacterial cells can be centrifuged away, giving a soup of phages, each of which carries on its surface the antibody encoded by the inserted gene. CAT has developed libraries that contain ten billion phage antibodies, says Kevin Johnson, the company's chief technology officer. Antibodies that bind to a particular target are then fished, or 'panned', from this soup, typically using a plastic surface on which the protein in question has been immobilized.

Capture molecules on a protein chip need to bind with high affinity because some of the most interesting proteins in a biological sample — such as hormones, growth factors and intracellular signalling proteins — are present only at very low concentrations. In practice, capture molecules must be able to identify target proteins from nanomolar ( $10^{-9}$  molar) to picomolar ( $10^{-12}$  molar) solutions.

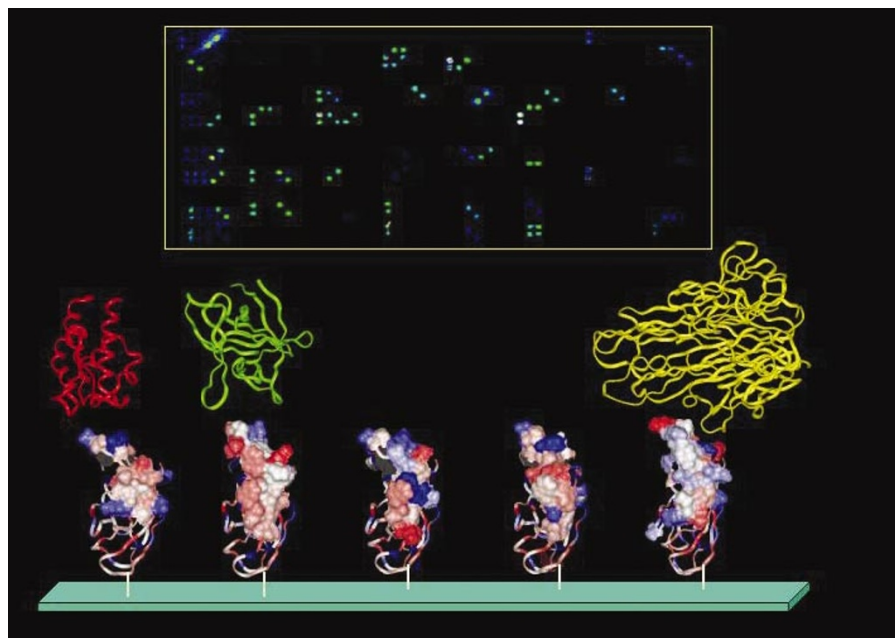
Success in identifying such molecules for large numbers of proteins is simply a matter of statistics — the bigger the library, the greater the chance of finding high-affinity antibodies. CAT has been successful in identi-



Turn up the volume: Kevin Johnson says CAT has created libraries of some ten billion antibodies.



Leigh Anderson of Large Scale Biology backs antibodies as the most reliable 'bait' molecules.



The protein-chip concept: capture molecules immobilized on a surface (bottom) bind to specific target proteins (here in red, green, yellow), giving a signal that can be read by viewing the chip (top).

fyng specific antibodies that bind with high affinity to individual targets — for example, it has an antibody against tumour-necrosis factor- $\alpha$  (ref. 2) that is now in advanced clinical trials as a treatment for rheumatoid arthritis. But routinely coming up with high-affinity antibodies for thousands of target proteins is a different matter. Bigger libraries are needed, but the cumbersome step of fermenting bacterial cultures in phage display makes it difficult to expand libraries further.

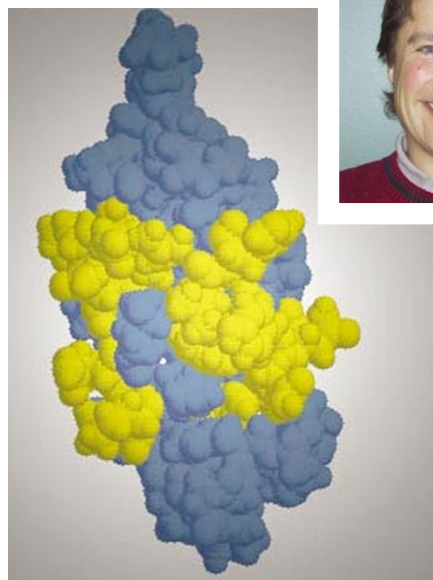
So CAT is now turning to a new ‘ribosome display’ technique developed by the Delaware-based company Aptein, which it purchased in 1998. Here, the phage is substituted with a ribosome, the cellular machinery that translates mRNA into protein. Normally, an mRNA molecule passes through the ribosome like ticker-tape and is released, along with the newly synthesized protein molecule, when a sequence of three bases known as a ‘stop codon’ is reached. In Aptein’s technology, stop codons are eliminated so that the completed antibody and its mRNA remain bound together on the ribosome. The system, which CAT is now optimizing, is entirely cell-free and so is more amenable to automation. This should make it possible to construct libraries that are orders of magnitude larger than those created using phage display.

But antibodies have their drawbacks. They tend to be denatured — lose their structure — when heated or exposed to other stresses such as changes in pH. And companies that are trying to enter the field are also restricted by that fact that many key antibodies and antibody technologies are covered by patents. So newcomers have been driven to look for alternatives.

One leading contender is ‘combinatorial

engineering’ of protein scaffolds. A scaffold is a domain of a large protein that, like an antibody, can be made to bind to an enormous range of other proteins by subtly changing its sequence of amino acids. The scaffolds are more stable than antibodies to heat or other stresses, and are sufficiently small that they, or the genetic sequences to encode them, can be synthesized chemically.

A handful of competing systems are being developed by various start-up firms. But two companies — Affibody and Phylos — are currently battling for supremacy in this area. Each claims that its proprietary scaffolds will allow the development of chips that can analyse tens of thousands of proteins simultaneously.



Stefan Ståhl (inset) says that ‘affibody’ scaffold domains can bind to a wide variety of proteins.



Phylos, based in Lexington, Maryland, uses a 100-amino-acid domain from a structural protein called human fibronectin for its scaffold. Like the binding sites of antibodies, this domain consists of a rigid unit that supports loops of varying lengths. Small changes in amino-acid sequence change the shape of these loops, and hence the structures to which the domain will bind. By substituting different amino acids in the loops held by a scaffold, trillions or more variants can be made. Libraries of these can then be searched to identify any that bind to a particular protein.

Phylos has created vast libraries of scaffold variants, which it calls trinetctins, using a proprietary mRNA-display system. This is a cell-free system similar to CAT’s ribosome display, except that the mRNA remains bound to the proteins produced and the ribosomes are washed away<sup>3</sup>. “It is a very simple process and easy to automate,” says Richard Wagner, Phylos’s head of R&D.

Using this system, Phylos produces libraries of some  $10^{13}$  variants. By applying several rounds of panning and washing away loosely bound proteins at each round, Wagner claims to have identified molecules that can capture their targets with nanomolecular or higher affinity — although nothing has yet been published. Targeted proteins include immune signalling molecules and their receptors.

### Winning combinations

Affibody, based in Stockholm, Sweden, uses as its scaffold a domain of staphylococcus protein A, a bacterial surface protein that consists of 58 amino acids<sup>4</sup>. It normally interacts, through a binding surface made up of 13 of its amino acids, with immunoglobulin G molecules — the main class of antibodies in the blood.

But by substituting these 13 amino acids, either singly or in combination, ‘affibody’ scaffolds can be made to bind to a wide variety of other proteins — in theory, it should be possible to create  $10^{16}$  different affibodies. And since 1998, using phage display, Affibody has developed libraries containing up to  $10^8$  variants.

“We have never failed to find a binder for the proteins we throw into our soup of variants,” says Stefan Ståhl, Affibody’s chief scientific officer. Again, the company conducts repeated rounds of panning, and has developed affibodies with nanomolar affinities for several proteins<sup>5</sup>.

Given the limitations of phage display, Affibody is looking for a cell-free alternative to boost library sizes. “We are working on three different selection systems which do not involve passage of phages through a bacterial system and which are more suited to automation,” says Ståhl. He declines to give

details, citing commercial confidentiality.

But not everyone is convinced that approaches that rely on *in vitro* screening for high-affinity binders will deliver the goods. Apart from the difficulty of generating sufficiently large libraries, another problem is that the selected molecules might cross-react with other proteins. "I think the body is a better vehicle for screening than a library," says Ian Humphery-Smith of the University of Utrecht in the Netherlands, who has founded a company called Glaucus Proteomics.

Humphery-Smith is instead deriving antibodies using mice engineered to have a human immune system. Each mouse is immunized with multiple antigen proteins, and after 40 days its blood plasma is screened against a protein chip with the same antigens attached. "In this way we can screen for cross-reactivity of antibodies as well as specificity and high affinity, in one go," says Humphery-Smith. The system will spawn a chip containing 150,000 antibodies within two-and-a-half years, he claims.



**Richard Wagner** says Phyllos's system is easy to automate.

Other companies have eschewed proteins as capture molecules, and have turned instead to 'aptamers' — short strings of DNA or RNA that constitute specific binding partners for a range of proteins. Work on aptamers is at an earlier stage than that with protein-based capture molecules, but if the technology can be made to work, it has strong appeal — aptamers can be synthesized chemically and the chips could be produced using the same high-throughput techniques already perfected for DNA microarrays.

"It's a fascinating idea," says Anderson of Large Scale Biology. But he warns that aptamers have not yet been shown to bind specifically to individual proteins when confronted with a complex mixture.

The leader in the aptamer field is SomaLogic of Boulder, Colorado, which is collaborating with a major player in proteomics, Celera Genomics of Rockville, Maryland. SomaLogic has produced a library of 10<sup>15</sup> DNA molecules in which thymidine — one of the four 'letters' of the genetic code — is replaced with bromodioxuracil. Already, the company has identified aptamers that bind with high affinity to a range of target proteins. "We expect to have a thousand proteins on a chip by next summer," says Larry Gold, SomaLogic's chief executive officer.

In SomaLogic's 'photoaptamer' system<sup>6</sup>, one end of each aptamer is covalently bound to the chip surface. Target proteins in a sample are captured by their individual aptamers, and the chip is then exposed to ultraviolet light, which causes the bromodi-

## Some companies are starting to think about who their customers will be, and what they will expect.

oxuracil to cross-link with the captured proteins. Unbound proteins can then be washed away and the remainder can be identified using a general protein stain.

This simple detection system would solve the second major technical challenge of protein chips — creating a reliable and rapid read-out system. General protein stains cannot be used in systems in which the capture molecules are also proteins. And, unfortunately, simply labelling all the proteins in a sample with a fluorescent tag, as with the mRNAs detected by conventional DNA microarrays, is not a viable option, because different proteins take up the tags to different extents.



**Larry Gold's** SomaLogic leads the aptamer field.

Currently, most systems rely on adding labelled antibodies to the proteins after they have been captured on the chips. But these 'sandwich assays' are hard to use as it is difficult to get sufficiently large numbers of antibodies into solution, and antibodies may bind to non-target proteins.

### Sandwich rapped

For this reason, companies working with protein-based capture molecules are trying to develop systems that do not involve sandwich assays. One idea is to label the capture molecules so that they will signal when they have bound to their target protein. Affibody, for instance, is working on a system dubbed FLAME, which is based on a technique called fluorescence resonance energy transfer. In this system, the affibody capture molecules are engineered to include two tags, one of which fluoresces but is suppressed by the second if it is close by. When an affibody binds to its target, the two tags move away from one another and fluorescence occurs<sup>7</sup>.

Other approaches rely on detecting the physicochemical changes that occur when a capture molecule binds to its target molecule. Several companies, including BiaCore of Uppsala, Sweden, are developing a method called surface plasmon resonance, which detects differences in refractive index at the surface of a capture molecule<sup>8</sup>. The Dutch electronics giant Philips, based in Eindhoven, is collaborating with Humphery-Smith to measure the changes in potential difference between the chip surface and the sample solu-

tion that accompany binding. It is also using microelectromechanical devices to record the physical changes associated with binding.

SomaLogic, meanwhile, says that it is refining its own simple detection method to cope with another problem — the fact that the concentrations of different proteins in a biological sample can vary over several orders of magnitude. Detecting them all may require samples to be split, diluted to different extents, and analysed repeatedly on duplicate chips — or, alternatively, splitting capture molecules for high- and low-concentration proteins between different chips. SomaLogic believes that it can solve the problem using single chips, but so far it has revealed no details.

As companies work to resolve the technical problems of making protein chips viable, some are also starting to think about who their customers will be, and what they will expect.



**Cell biologist** Gavin MacBeath will use the chips.

"There really has been too little concentration on what the customers want, and too little consideration of what they will be prepared to pay," says Johnson of CAT.

Unless there is a significant technical advance, argues Johnson, the cost of protein chips will increase geometrically with the number of proteins on each chip. This is why CAT has, for now, opted for low-density chips, each addressing custom-made subsets of proteins.

Such chips will find their uses — in basic biology as well as diagnostics. "A chip that could follow a few dozen proteins in a cell-signalling pathway would be a great way to unravel the whole pathway," says Gavin MacBeath of Harvard University's Bauer Center for Genomics Research. "We could see which proteins upregulate in compensation when a particular target protein, like a receptor, is blocked."

As for which — if any — technology can provide the genuine proteomics breakthrough that would give rise to chips carrying tens of thousands of protein-capture molecules, investors are now placing their bets. ■

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

1. Vaughan, T. J. *et al. Nature Biotechnol.* **14**, 309–314 (1996).
2. Mahler, S. M., Marquis, C. P., Brown, G., Roberts, A. & Hoogenboom, H. R. *Immunotechnology* **3**, 31–43 (1997).
3. Roberts, R. W. *Curr. Opin. Chem. Biol.* **3**, 268–273 (1999).
4. Nord, K. *et al. Nature Biotechnol.* **15**, 772–777 (1997).
5. Nord, K. *et al. Eur. J. Biochem.* **268**, 4269–4277 (2001).
6. Brody, E. & Gold, L. *Rev. Mol. Biotechnol.* **74**, 5–13 (2000).
7. Karlstöm, A. & Nygren, P.-A. *Anal. Biochem.* **295**, 22–30 (2001).
8. Nelson, R. W., Nedelkov, D. & Tubbs, K. A. *Electrophoresis* **21**, 1155–1163 (2000).

CAT ♦ [www.cambridgeantibody.com](http://www.cambridgeantibody.com)

Phyllos ♦ [www.phylos.com](http://www.phylos.com)

Affibody ♦ [www.affibody.com](http://www.affibody.com)

Glaucus Proteomics ♦ [www.glaucusprot.com](http://www.glaucusprot.com)

SomaLogic ♦ [www.somallogic.com](http://www.somallogic.com)

BiaCore ♦ [www.biacore.com](http://www.biacore.com)