Biochemistry's new look

Until now, metabolomics researchers have had to adapt technology developed mainly for proteomics. But there are now solutions designed with them in mind. Nathan Blow reports.

Metabolomics — the comprehensive study of metabolic reactions — is gaining ground alongside its older siblings genomics and proteomics. "Unlike some of the other 'omics' that we have seen, metabolomics is going to produce a lot of useful information right from the start," says Gary Siuzdak, professor of molecular biology at the Scripps Research Institute in La Jolla, California. He is one of a growing number of biologists using advanced technology to explore biochemical questions on a scale that would have seemed impossible a decade ago.

"The metabolome is the best indicator of an organism's phenotype," says David Wishart at the University of Alberta in Edmonton, Canada. Wishart was one of the instigators of the Human Metabolome Project, a US$7.5-million effort funded by Genome Canada to systematically characterize the metabolites of the human body. He gives the example of a person holding their breath for five minutes. Although genomic or proteomic analysis would not provide any evidence of stress during this short period — even as the person turns blue — metabolite profiles would show dramatic changes within the body.

Unlike a genome or even a proteome, however, a metabolome is tricky to pin down. Wishart notes that although researchers know there are 3 billion base pairs in the human genome, if you ask biochemists how many small-molecule metabolites there are in the human body, they come back with numbers ranging from 3,000 to 100,000. And this poses a real challenge for metabolomics research, as both ends of the scale could be correct.

The Human Metabolome Project has pegged the number of endogenous metabolites in the human body at around 3,000 — which most researchers agree on. But humans also take in small molecules from the environment — preservatives in food, chemicals in the air, metabolites produced through the breakdown of drugs and toxins — making an exact figure hard to determine.

Separation anxiety

With metabolites in such a state of flux, researchers do not have an easy task. Nevertheless, advances in chromatography, mass spectrometry (MS) and nuclear magnetic resonance spectroscopy (NMR) are allowing them to make headway in defining different

LECO

The Pegasus 4D GCxGC MS TOF system enables multidimensional approaches to GC separation.

Gary Siuzdak: developing new approaches to metabolite identification.

metabolomes and understanding how changes in the concentrations of metabolites relate to human health and disease.

One of the problems is that metabolites come in a variety of chemical forms. "I would say one of the real challenges of metabolomics is that each metabolite is its own unique puzzle," says Trent Northen, a scientist at Lawrence Berkeley National Laboratory in California. And in most cases the first step in solving the puzzle is isolating the metabolite for analysis.

No one separation method works for all metabolites, so researchers rely on combinations of gas chromatography (GC), liquid chromatography (LC) and emerging capillary electrophoresis (CE).

Historically, GC separation has had the edge. "GC–MS technology may not be sexy, but huge databases are available," says Northen. These GC–MS databases, compiled over more than four decades, enable researchers to compare a wide range of spectra to arrive at a chemical identification.

Multidimensional GC, often called 'GC×GC' or two-dimensional GC, offers even better separation. "When people are doing GC×GC, they are trying to get more separation chroma-

tographically of very complex samples," says Steven Fischer, a senior applications chemist at Agilent Technologies in Santa Clara, California. To achieve this, GC×GC uses two separation phases, such as a non-polar and a polar phase, in two capillary columns in series in the instrument.

Agilent recently introduced the 7890 GC system, which can perform multidimensional GC, and Thermo Fisher Scientific in Waltham Massachusetts, has developed the Trace GC×GC system. The Pegasus 4D GC×GC MS time-of-flight (TOF) system from LECO, based in St Joseph, Michigan, uses a thermal modulator placed between the two GC columns to collect effluent from the first column before going into the second phase of separation. The power of the multidimensional approach is starting to be reported. In May this year a group reported the use of GC×GC with LC–MS to generate a draft metabolic network for the single-celled alga *Chlamydomonas reinhardtii*[1].

Class action

GC is particularly useful for mixtures of volatiles, such as steroids, saccharides and sugar alcohols, which can be sent directly into the gas phase for separation. Metabolites in human biofluids and tissues therefore present a technical challenge for GC, as most are not volatile. Non-volatile metabolites either need complicated chemical transformation before GC or

separation by other types of chromatography.

One of these is high-performance liquid chromatography (HPLC), a well-established lab workhorse. It uses a combination of solvents, pressure and matrix particle sizes to separate molecules on the basis of their retention times in a column packed with matrix. HPLC can separate a broad range of metabolites, including non-volatiles, and remains a favourite among metabolomics researchers.

Most advances in HPLC involve increases in the pressure applied and changes in matrix particle size. Ultra-performance liquid chromatography (UPLC), commercialized by Waters Corporation in Milford, Massachusetts, is becoming more widely used in the metabolomics community. It takes advantage of higher pressure (83 megapascals compared with 21 megapascals for HPLC) and smaller particles (less than 2 micrometres diameter compared with 3 micrometres for HPLC) to obtain faster separation times.

But like GC, HPLC has technical stumbling blocks. Reversed-phase HPLC (in which the stationary phase is non-polar) is often used for metabolomics analysis, but reversed-phase separation often fails with hydrophilic metabolites. These tend to be so water soluble that they interact poorly with the non-polar bonding phase and are

rapidly eluted, according to Phil Koerner, a senior technical manager from chromatography specialists Phenomenex in Torrance, California.

So in 2007, Phenomenex introduced the Luna HILIC column. "I like to refer to it as reverse reverse-phase chromatography," says Koerner. In the HILIC approach, the weak solvent, which is applied first, is a polar organic solvent (not water as in reversed-phase HPLC), and the strong solvent, applied second, is water. This causes the order of elution to be completely reversed, with the most hydrophilic compounds being eluted last. Although Koerner acknowledges that the HILIC approach is not new, it was the need to separate hydrophilic metabolites on the large scale required by metabolomics that led Phenomenex and other companies, such as Waters and Tosoh Bioscience of Stuttgart, Ger-

many, to start supplying a greater number and range of HILIC columns.

Capillary electrophoresis followed by MS (CE-MS) is not yet so popular with the metabolomics community as either GC or HPLC, but several developers are hoping to change this. "It can be very difficult to use this approach," acknowledges Ryuji Kanno, president of Human Metabolome Technologies based in Tokyo, Japan. This approach uses electrophoretic mobility to separate low-molecular-weight ionic compounds that are difficult to separate by GC or HPLC. The company has been working closely with Agilent to develop optimized reagents and capillary columns, and is providing training with Agilent's CE-qTOF MS system to make the CE approach more accessible to metabolomics researchers, says Kanno.

Mass spectrometry is not the only method that can be used to detect metabolites once separated. Wishart and his colleagues recently compared MS and NMR to look at metabolites in cerebral spinal fluid[2]. They found little overlap in the metabolites detected by the two methods, and the conclusion was clear: "We do not have a single perfect metabolite detector," says Wishart.

MS and NMR each have their supporters. "One of the main strengths of NMR is that

HILIC columns are making the hunt for hydrophilic metabolites easier.
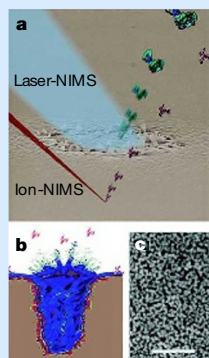
PHENOMENEX

DARK MATTER

"There are a lot of small molecules that we do not even know about yet," says Arthur Castle, programme director for the Roadmap Metabolomics Technology development programme at the US National Institutes of Health in Bethesda, Maryland. Metabolomics has a good handle on analysing human primary metabolites, but when it comes to lipids, secondary metabolites, xenobiotics and the products of gut microflora, we are just scratching the surface, says Castle.

The problem is part technological, part informatics. Steven Fischer at Agilent Technologies in Santa Clara, California, points out that some compounds are not stable and undergo chemical transformation during separation. Researchers will probably still see these transformed molecules by mass spectrometry (MS), but they may

be misidentified, highlighting the need for follow-up experiments. And this is where metabolomics has an advantage over the other 'omics'. "There is so much knowledge of biochemistry that when we find a potential biomarker or a new drug mechanism we already know a lot about it," says Michael Milburn, chief scientific officer at Metabolon in Durham, North Carolina.

Trent Northen, now at the Lawrence Berkeley Laboratory in Berkeley, California, and Oscar Yanes, working in Gary Siuzdak's lab at the Scripps Research Institute in La Jolla, California, may have developed a new way to get at some of these 'unknown' molecules with

an ionization technique called nanostructure-initiator mass spectrometry (NIMS)[4]. The idea is to transfer a biomolecule into the gas phase from a nanostructured surface simply by making that surface disappear. "We came up with the idea of putting a wax underneath, so when the nanostructured surface was irradiated it would melt and vaporize, allowing the molecule to go into the gas phase," says Northen. They finally came up with a perfluorinate surface for the trapped initiator phase. Siuzdak's group found that NIMS worked not only for proteins, but also for small molecules such as metabolites. "Perfluorinates do not ionize well, so it allows us to

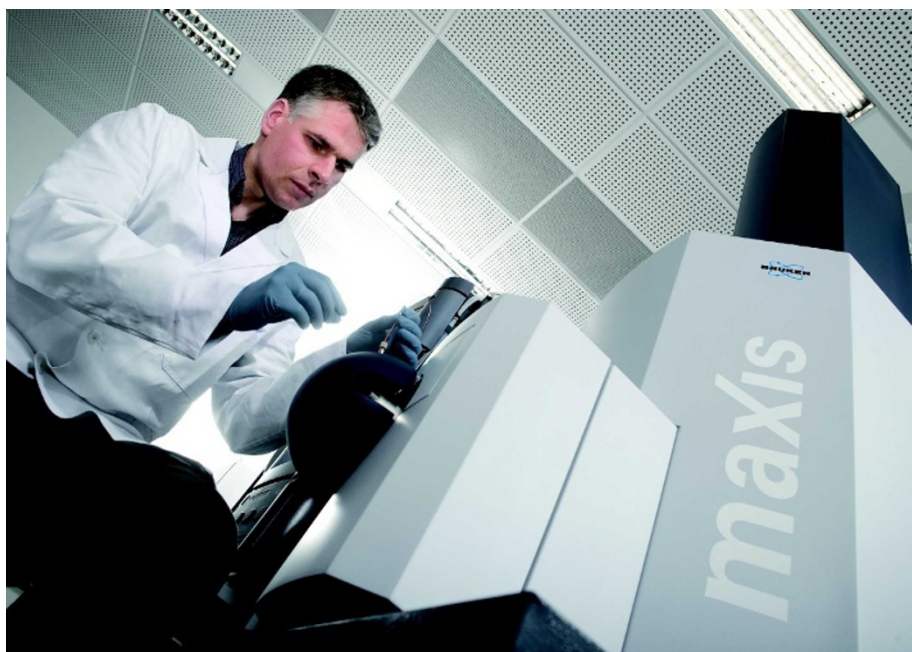*NIMS: nanostructured surface releases material.*

REF. 4

see things in the lower mass region where metabolites like to hang out," says Siuzdak.

Using perfluorinated materials may have another advantage for metabolomics. "As these are highly hydrophobic surfaces, we can apply very dirty complex samples, such as blood, for direct analysis," says Yanes. If you put a drop of urine or blood on the NIMS chip the metabolites will attach, but all the salts and other chemicals that normally interfere with MS stay in solution. Yanes is now exploiting this property to follow drug metabolism by looking at uptake in blood, clearance in urine and tissue localization.

For Siuzdak, exploring the 'unknown' metabolite world is an exciting prospect. "We are getting involved in an area where we don't know what the molecules' structures are or what they do, so it is really just a fantastic area for discovery." N.B.

The maXis system from Bruker Daltonics can use both UPLC and CE separation approaches.

it is an unbiased, universal detector," says Jack Newton, a product manager at Chenomx in Edmonton, Canada, which was co-founded by Wishart in 2000. This attribute, along with NMR's ability to determine structure and perform quantitative analysis is particularly attractive to metabolomics researchers who need a way to compare and exchange results between labs. "The move is afoot — people want to get to that common language of compound names and concentrations," says Wishart, as this will make integrating data sets and obtaining systems-level views of cell physiology possible.

The challenge with NMR is instrument sensitivity — NMR is less sensitive than MS, often identifying far fewer metabolites in the same sample. "For us, the relevant question is how sensitive do you need to be," says Newton. He says researchers at Chenomx have performed many studies in which biologically meaningful differences between samples were easily captured with NMR, even though some compounds in the samples probably fell below the sensitivity limits of the instrument (see 'Dark matter').

MS, on the other hand, is a very sensitive method for metabolite identification and, unlike NMR, is easily coupled to upstream separation techniques. Siuzdak says his group can see thousands of molecules in an MS analysis — and that number can be doubled by changing from positive- to negative-ion mode. And by using both reversed-phase chromatography and HILIC columns, they are seeing more hydrophilic compounds in their analyses than before. "I would venture that we are now seeing over an order of magnitude more than what you would see with NMR," he says.

Detector development

As researchers in the MS camp turn towards TOF and ion-trap MS instruments for metabolite analysis, developers are responding to their complex needs. Bruker Daltonics in Billerica, Massachusetts, has introduced the maXis ultra-high resolution (UHR)-TOF MS system, which can accommodate both UPLC and CE separation. Applied Biosystems in Foster City, California, in collaboration with MDS SCIEX in Toronto, Ontario, have the ion-trap system 4000 QTrap LC/MS/MS that can interface with Applied Biosystem's LightSight software for small-molecule analysis and identification.

WINE-OMICS

For Kirsten Skogerson at the University of California, Davis, wondering about how chemical composition affects the flavour and body of a wine took her from a degree in viticulture and enology into metabolomics research. When Skogerson arrived in Oliver Fiehn's lab as a postgrad she looked for a project that would marry Fiehn's expertise in metabolomics and her interest in wine.

"There are so many questions in wine science that you could start to answer by doing a global analysis," she says. A deeper understanding of the biochemistry of grape-juice fermentation could help the winemaking industry by complementing the arts of the traditional wine taster. So Skogerson and Fiehn set out to survey wine 'metabolomes', in search of key chemical components contributing to body.

Using proton nuclear magnetic resonance (NMR) and gas chromatography–mass spectrometry (GC–MS), they looked at 17 different white wines with a wide range of body. For GC–MS analysis, they first removed the alcohol under reduced pressure and then ran samples on a LECO Pegasus IV GC TOF MS system and analysed the spectra using the BinBase program developed in Fiehn's lab. Each wine was also directly analysed on a Bruker Daltonics 600 mHZ NMR instrument with the resulting peaks being compared to the commercially available Chenomx NMR database for metabolite identification. "When you think about it, you have the grape metabolome being acted on by the yeast, plus the added complexity from the yeast metabolome, so the metabolite profile of a wine is very complex," says Skogerson.

They found a total of 413 metabolites among the wines — probably only a small fraction of the wine metabolome — of which 108 could be positively identified. And in both data sets, the amino acid proline showed a positive correlation with body as assessed by trained wine tasters. How proline relates to body is not yet clear, however. "That is the hard part of being in metabolomics — you get clues, but the follow-up is the real challenge," says Skogerson. Still, she thinks proline could be used as marker for a wine's viscosity.

Red-wine drinkers have not been forgotten. Bruker Daltonics in Billerica, Maryland, has profiled red wines for important polyphenolic secondary metabolites such as tannins, flavonoids and anthocyanins. This demonstration used the Acquity ultra-performance liquid chromatography system from Waters to separate red wine metabolites for analysis by Bruker's LC-ESI QTOF MS instrument as well as analysis by NMR coupled with Bruker's BioSpin Spectral Base analysis package.

Does knowing the chemistry behind that wonderful bottle of wine take away from the pleasure? Not according to Skogerson. "Science has the potential to bring the art of winemaking to a higher level."   N.B.

699
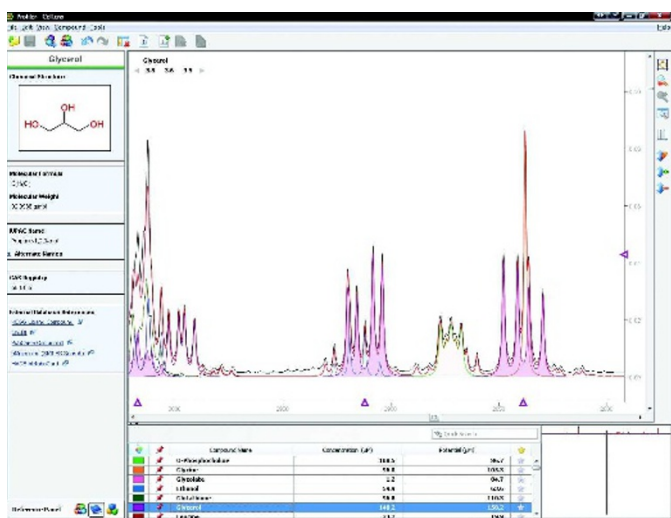
**Chenomx has developed a searchable NMR database for metabolomics.**

Both Agilent and Thermo Fisher Scientific also offer MS systems and software packages designed for metabolite analysis.

Some researchers and developers are designing platforms to bring the two camps closer together — incorporating NMR and MS instruments in a single system. Bruker BioSpin in Billerica, Massachusetts, has developed the Metabolic Profiler, a system that combines a liquid handler, the Avance III NMR spectrometer and an LC-electrospray ionization (ESI)-microTOF MS, all under the control of a single data-management and analysis system.

But what researchers dream of is a single detection 'chip' for all metabolites. "In my lab we have four platforms, and each platform looks at a certain part of the metabolome," says Oliver Fiehn, a metabolomics researcher at the University of California, Davis. But he doubts that a single chip could ever become reality.

"The lack of such a technology is the Achilles heel of metabolomics," says Wishart, noting that the most that researchers can analyse at any one time with current technologies is 10–15% of the entire metabolome — and even that's stretching it.

"The big bottleneck is really compound identification," says Fiehn. Unblocking it will need the addition of many more well-annotated reference spectra in the databases.
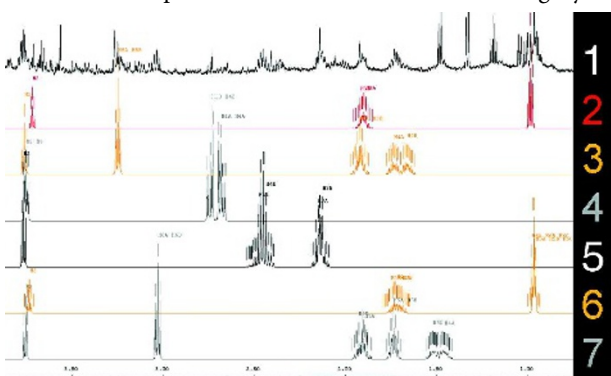
Comparisons of samples to reference spectra databases help reveal the identity of metabolites.

And that will take time. Chenomx was founded with the aim of developing a database for NMR analysis, and that has taken several years of intensive effort, says Newton. Different chemical environments can influence a compound's NMR spectra, so researchers at Chenomx had to acquire spectra at ten pHs, ranging from 4 to 9, for each of the more than 300 reference compounds now in their proprietary database.

Metabolite databases for MS have also been springing up as more researchers move into the field. One of the first was METLIN (http://metlin.scripps.edu), a publicly accessible database that was started in Siuzdak's lab. "We currently have 23,000 metabolites in there," says Siuzdak, of which around 2,500 are identified endogenous metabolites. METLIN also contains a set of about 8,000 theoretical di- and tripeptides along with theoretical lipids, drugs and metabolites.

To expand the scope of METLIN, Agilent has collaborated with the Scripps Center for Mass Spectrometry to analyse chromatographic standards and add information about mass and retention time, with the intent of using these properties in addition to isotope pattern matching for identification. "Our goal is to get to the point where the most common metabolites encountered by researchers are easily identifiable," says Agilent's Fischer.

Gregory Stephanopoulos, a chemical engineer at the Massachusetts Institute of Technology in Cambridge, is taking a different approach to metabolite identification. Several years ago a student approached him with an interesting metabolomics project, but the catch was that the lab would first have to increase the number of reference spectra in its library to enable metabolite identification. Although Stephanopoulos liked the project, he did not like the idea of simply collecting spectra to fill a database. "I thought that there had to be a better way to deal with the issue," he recalls.
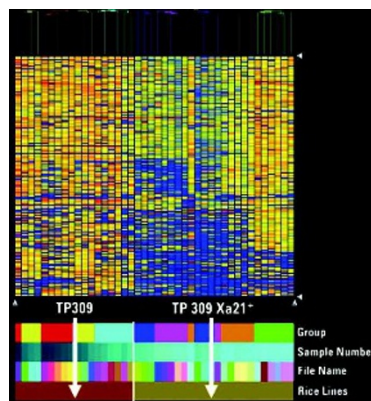
The result is a web-based program called SpectConnect, which was launched in 2007 to help researchers identify important metabolites that might serve as biomarkers[3]. The SpectConnect algorithm tracks and catalogues GC–MS spectra that are conserved in multiple samples — an indication that these represent real compounds instead of noise in the sample, helping to guide researchers with their follow-up efforts at full metabolite identification.

Numbers game

The good news for metabolomics researchers is that NMR and MS metabolite databases are increasing in both number and size as new metabolomes are analysed. "One of the things that changed for us over the past 18 months is the places we are applying the technology," says Michael Milburn, chief scientific officer at Metabolon in Durham, North Carolina. Metabolomic approaches are now addressing biological questions in areas ranging from drug discovery and cosmetics development to plant science and winemaking (see 'Wine-omics'). Publicly accessible databases include MassBank for high-resolution ESI mass spectra of metabolites (www.massbank.jp), BinBase for processing and analysing of dissimilar MS spectra (http://sourceforge.net/projects/binbase), and MetWare (http://msbi.ipb-halle.de/msbi/metware) for the storage and analysis of metabolomic experiments. Commercial databases include Metabolon's, containing spectra of more than 6,000 reference metabolites, and Bio-Rad's KnowITAll spectral database of more than 1.3 million entries, including MS and NMR references.

But there's still a way to go before metabolite identification is as simple as 'query and get a chemical name'. "The database changes have been encouraging," say Stephanopoulos. But not enough to change his mind about the need for tools such as SpectConnect.

Arthur Castle, programme director for the Roadmap Metabolomics Technology development programme at the US National Institutes of Health, has seen the pieces falling into place over the past couple of years. "The technology is very close to being there — it is just a question of putting it all together now," he says. ■

Nathan Blow is technology editor for *Nature* and *Nature Methods*.

Samples can be grouped by the similarity of mass abundance profiles.

1. May, P. *et al. Genetics* **179,** 157–166 (2008).
2. Wishart, D. S. *et al. J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **15,** 164–173 (2008).
3. Styczynski, M. P. *et al. Anal. Chem.* **79,** 966–973 (2007).
4. Northen, T. R. *et al. Nature* **449,** 1033–1036 (2007).