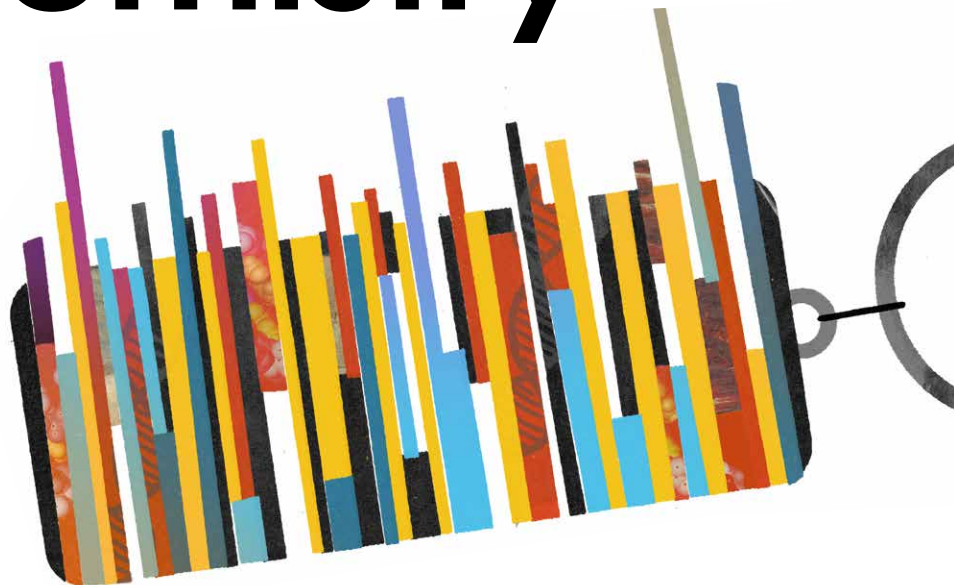


Unlocking the Chemistry of Life

by Jessica Stoller-Conrad



In just the span of an average lifetime, science has made leaps and bounds in our understanding of the human genome and its role in heredity and health—from the first insights about DNA structure in the 1950s to the rapid, inexpensive sequencing technologies of today. The 20,000 genes of the human genome—each made of a sequence of DNA bases—encode proteins to carry out the countless functions that are key to our existence. But we know much less about how this collection of proteins supports the essential functions of life than we do about the genes that built them.

In order to understand the role each of these proteins plays in human health—and what goes wrong within them to cause disease—biologists need to figure out what these proteins are and how they function. Several

decades ago, biologists realized that to answer these questions on the scale of the thousands of proteins in the human body, they would have to leave the comfort of their own discipline to get some help from a standard analytical-chemistry technique: mass spectrometry. Since 2006, Caltech's Proteome Exploration Laboratory (PEL) has been building on this approach to bridge the gap between biology and chemistry, in the process unlocking important insights about how the human body works.

Scientists can easily sequence an entire genome in just a day or two, but sequencing a proteome—all of the proteins encoded by a genome—is a much greater challenge, says Ray Deshaies, protein biologist and founder of the PEL. "One challenge is the amount of protein. If you want to sequence a person's DNA from a few of their

cheek cells, you first amplify—or make copies of—the DNA so that you'll have a lot of it to analyze. However, there is no such thing as protein amplification," Deshaies says. "The number of protein molecules in the cells that you have is the number that you have, so you must use a very sensitive technique to identify those very few molecules."

The best means available for doing this today is called shotgun mass spectrometry, Deshaies says. In general, mass spectrometry allows researchers to identify the amount and types of molecules that are present in a biological sample by separating and analyzing the molecules as gas ions, based on mass and charge; shotgun mass spectrometry—a combination of several techniques—applies this separation process specifically to digested, broken-down proteins, allowing researchers to identify the types and

amounts of proteins that are present in a heterogeneous mixture.

The first step of shotgun mass spectroscopy entails digesting a mixture of proteins into smaller fragments called peptides. The peptides are separated based on their physical properties, and then they are sprayed into a mass spectrometer and blasted apart via collisions with gas molecules such as helium or nitrogen—a process

of this field-changing technology, he and colleague Barbara Wold, the Bren Professor of Molecular Biology, applied for and received a Department of Energy grant for their very own mass spectrometer. When the instrument arrived on campus, demand began to surge.

“Barbara and I were first just doing experiments for our own labs, but then other people on campus wanted us to

that gap as director of the PEL.

Hess, who came from a proteomics lab at the National Institutes of Health, knew the challenges of running an interdisciplinary center such as the PEL. Although the field of proteomics holds great promise for understanding big questions in many fields, including biology and medicine, mass spectrometry is still a highly technical method involving analytical chemistry and data science—and it’s a technique that many biologists were never trained in. Conversely, many chemists and mass spectrometry technicians don’t necessarily understand how to apply the technique to biological processes.

By encouraging dialogue between these two groups, Hess says, the PEL crosses that barrier, helping apply mass spectrometry techniques to diverse research questions from more than 20 laboratories on campus.

Creating this interdisciplinary and resource-rich environment has enabled a wide breadth of discoveries, says Hess. One major user of the PEL, chemist David Tirrell, has turned to the center for many collaborations involving a technique he developed with former colleagues Erin Schuman and Daniela Dienerich called BONCAT (for “bioorthogonal noncanonical amino-acid tagging”).

BONCAT uses synthetic molecules that are not normally found in proteins in nature, and that carry particular chemical tags. When these artificial amino acids are incubated with certain cells, they are taken up by the cells and incorporated into all newly formed proteins in those cells. The tags then allow researchers to identify and pull out proteins from the cells, thus enabling them to wash away all of the other untagged proteins from other cells that aren’t of interest. When this method is combined with mass spectrometry techniques, it enables researchers to achieve specificity in their results and determine which proteins are produced in a particular subset of cells during a particular time.

that creates a unique fragmentation pattern for each peptide. This pattern, or “fingerprint,” of each peptide’s fragmentation can then be searched on a database and used to identify the protein the peptide came from.

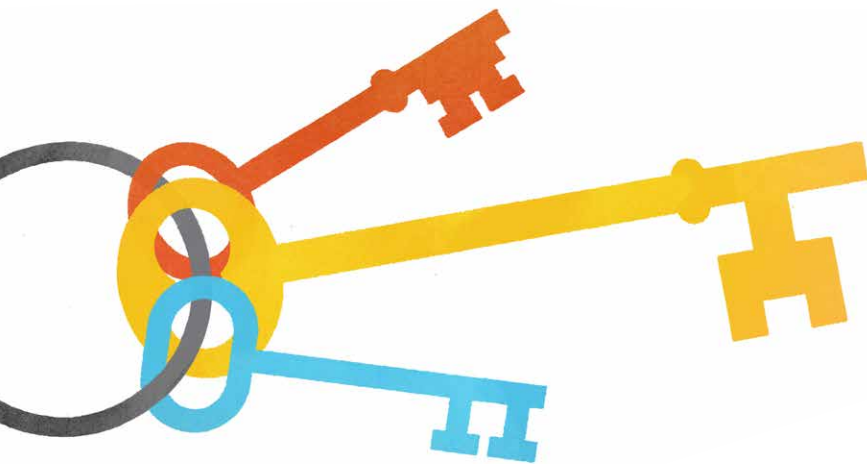
“Up until this technique was invented, people had to take a mixture of proteins, run a current through a polyacrylamide gel to separate the proteins by size, stain the proteins, and then physically cut the stained bands out of the gel to have each individual protein species sequenced,” says Deshaies. “But mass spectrometry technology has gotten so good that we can now cast a broader net by sequencing everything, then use data analysis to figure out what specific information is of interest after the dust settles down.”

Deshaies began using this shotgun mass spectrometry in the late 1990s, but because the technology was still very new, all of the protein analysis had to be done at the outside laboratories that were inventing the methodology. In 2001, after realizing the potential

help them apply this technology to their research problems,” Deshaies says.

So he and Wold campaigned for a larger, ongoing center where anyone could begin using mass spectrometry resources for protein research. In 2006, Deshaies and then chair of the Division of Biology (now the Division of Biology and Biological Engineering) Elliot Meyerowitz petitioned the Gordon and Betty Moore Foundation to secure funding for a formal Proteome Exploration Laboratory, as part of the foundation’s commitment to Caltech.

The influx of cash dramatically expanded the capabilities and resources that were available to the PEL, allowing it to purchase the best and fastest mass spectrometry instruments available. But just as importantly, it also meant that the PEL could expand its human resources, Deshaies adds. It was mostly students who were running the instruments in the Deshaies lab, he says, so when they graduated or moved on, gaps were left in expertise. Sonja Hess came to Caltech in 2007 to fill



“In my own laboratory, we work at making sure the method is adapted appropriately to the specifics of a biological problem. But we rely on collaborations with other laboratories to help us understand what the demands on the method are and what kinds of questions would be interesting to people in those fields,” Tirrell says.

For example, Tirrell collaborated with biologist Paul Sternberg and the PEL, using BONCAT and mass spectrometry to analyze specific proteins in a few cells from a whole organism, a feat that had never been accomplished before. Using the nematode *C. elegans*, Sternberg and his team applied the BONCAT technique to tag proteins in the 20 cells of the worm’s pharynx and then used the PEL resources to analyze proteome-wide information from just those 20 cells. The results, including identification of proteins that were not previously associated with the pharynx, were published in the *Proceedings of the National Academy of Sciences* in 2014.

The team is now trying to use the same technique on a single pair of neurons that help the worm to sense and avoid harmful chemicals—a first step in learning which proteins are

essential to producing this responsive behavior. But analyzing protein information from just two cells is a difficult experiment, says Tirrell.

“The challenge comes in separating out the proteins that are made in those two cells from the proteins in the rest of the hundreds of cells in the worm’s body. You’re only interested in two cells, but to get the proteins from those two cells, you’re essentially trying to wash away everything else—about 500 times as much ‘junk’ protein

This next step is an important one, but Tirrell says that an advantage of the PEL is that the laboratory’s staff can focus on optimizing the very technical mass spectrometry aspects of an experiment, while researchers using the PEL can focus more holistically on the question they’re trying to answer. This was also true for biologist Mitch Guttman, who asked the laboratory to help him develop a mass spectrometry-based technique for identifying the proteins that hitchhike on a class

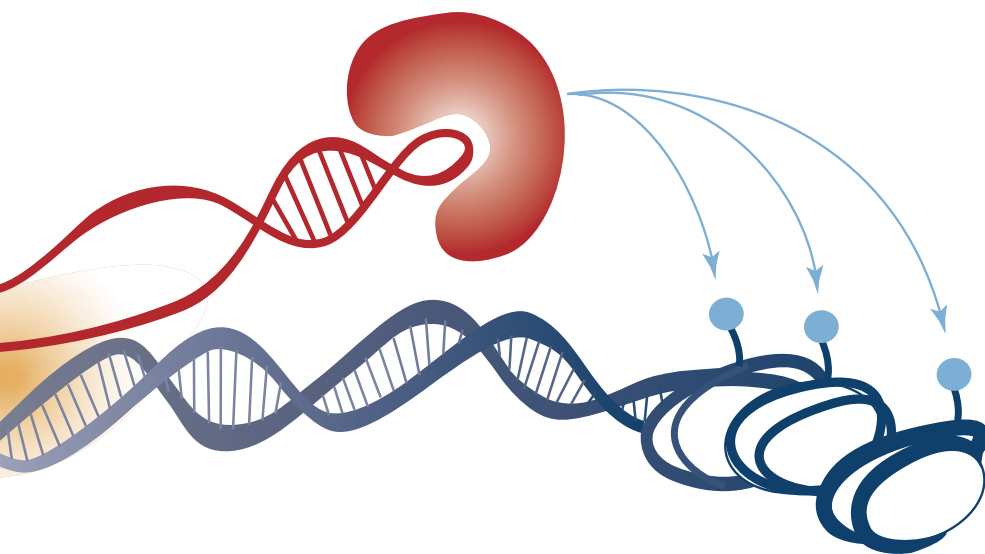
Although Deshaies says that the PEL resources have become invaluable to his work, he adds that what makes the laboratory unique is how it benefits the entire Institute—a factor that he hopes will encourage further support for its mission.

as the protein that you’re really interested in,” he says. “We’re working on these separation methods now because the ultimate experiment would be to find a way to use BONCAT and mass spec to pull out proteomic information from a single cell in an animal.”

of RNA genes called lncRNAs.

Long noncoding RNAs—or lncRNAs (pronounced “link RNAs”) for short—are abundant in the human genome, but scientists know very little about how they work or what they do. Although it’s known that protein-coding genes start out as DNA, which is transcribed into RNA, which is then translated into the gene product, a protein, lncRNAs are never translated into proteins. Instead, they’re thought to act as scaffolds, corraling important proteins and bringing them to where they’re needed in the cell. In a study published in April 2015 in *Nature*, Guttman used a specific example of a lncRNA, a gene called Xist, to learn more about these hitchhiking proteins.

“The big challenge to doing this was technical; we’ve never had a way to identify which proteins are actually interacting with a lncRNA molecule. By working with the PEL, we were able to develop a method based on mass spectrometry to actually purify and identify this complex of proteins interacting with a lncRNA in living cells,” Guttman says. “Once we had that information, we could really start to ask ourselves questions about these



The lncRNAs that Mitch Guttman studies don't encode proteins, rather they act as scaffolds, shuttling proteins to their destination within the cell. Above is a schematic of a lncRNA (red helical structure) in action, escorting a protein (red "C" shape) to DNA targets (light blue) within the cell.

proteins and how are they working.”

Using this new method, called RNA antisense purification with mass spectrometry (RAP-MS), Guttman’s lab determined that 10 proteins associate with the lncRNA Xist, and that three of those 10 are essential to the gene’s function—inactivating the second X chromosome in women, a necessary process that, if interrupted, results in the death of female embryos early in development.

Guttman’s findings marked the first time that anyone had uncovered the detailed mechanism of action for a lncRNA gene. For decades, other research groups had been trying to solve this problem; however, the collaborative development of RAP-MS in the PEL provided the missing piece.

Even Deshaies, who began doing shotgun mass spectrometry experiments in his own laboratory, now exclusively uses the PEL’s resources and says that the laboratory has played an essential support role in his work.

He studies the normal balance of proteins in a cell and how this balance changes during disease. In a 2013 study published in *Cell*, his laboratory focused on a dynamic network of protein complexes called SCF complexes, which go through cycles of assembly and dissociation in a cell depending on when they are needed.

Because there was no insight into how these complexes form and disassemble, Deshaies and his colleagues used the PEL to quantitatively monitor how this protein network’s dynamics were changing within cells. They determined that SCF complexes are normally very stable, but in the presence of a protein called Cand1 they become very dynamic and rapidly exchange subunits. Since some components of the SCF complex have been implicated in the development of human diseases such as cancers, work is now being done to see if Cand1 holds promise as a target for a cancer therapeutic.

Although Deshaies says that the PEL resources have become invaluable to his work, he adds that what makes the laboratory unique is how it benefits

the entire Institute—a factor that he hopes will encourage further support for its mission.

“The value of the PEL is not just about what it contributes to my lab or to Dave Tirrell’s lab or to anyone else’s,” he says. “It’s about the breadth of PEL’s impact—the 20 or so labs that are bringing in samples and using this operation every year to do important work, like solving the mechanism of X-chromosome inactivation in females.” *E&S*

Raymond Deshaies is a professor of biology, an investigator with the Howard Hughes Medical Institute (HHMI), and the executive officer for molecular biology at Caltech. His work is funded by HHMI and the National Institutes of Health (NIH).

Sonja Hess is the director of the Proteome Exploration Laboratory at Caltech. Three of the PEL’s instruments are funded by NIH, the Gordon and Betty Moore Foundation, HHMI, and the Beckman Institute. The laboratory is led by executive director Shu-ou Shan, professor of chemistry.

David Tirrell is Ross McCollum–William H. Corcoran Professor and professor of chemistry and chemical engineering and director of the Beckman Institute at Caltech. His collaboration with Paul Sternberg, the Thomas Hunt Morgan Professor of Biology and an investigator with HHMI, was supported by HHMI.

Mitchell Guttman is an assistant professor of biology at Caltech. His research with Xist was made possible by NIH.

