

OCTOBER 16, 2003

Researchers Create New Atlas of Yeast Proteins

Using high-tech robots and old-fashioned hard labor, Howard Hughes Medical Institute researchers have measured the abundance and pinpointed the cellular locations of more than 4,000 proteins in yeast.

Proteins are the workhorse molecules of the cell. They catalyze reactions, transport molecules within the cell and switch genes on and off. Measuring the abundance of and identifying the cellular locations of yeast proteins will be invaluable in helping to understand the complex biology of a relatively simple organism. Beyond that, however, the effort is emblematic of a shift in biological research, toward understanding how changes in the “proteome”—the interacting global network of proteins in a cell—can influence cellular “behavior.”

Research teams led by HHMI investigators [Erin K. O'Shea](#) and [Jonathan S. Weissman](#), both of the University of California at San Francisco, published two research articles describing their work in the October 16, 2003, issue of the journal *Nature*.

"Now, investigators can measure the abundance of proteins and follow their location with a degree of sensitivity that was never possible for any proteome in any organism."

- Jonathan S. Weissman

“We have now made the yeast proteome accessible in a way it simply wasn't before,” said Weissman. “Now, investigators can measure the abundance of proteins and follow their location with a degree of sensitivity that was never possible for any proteome in any organism. We believe that this capability really strengthens the status of yeast as the premier organism for the systems-biology approach to a coherent, comprehensive understanding of how the cell works.”

According to the researchers, determining the relative abundance of proteins in yeast will offer more insight into protein function than do studies of levels of messenger RNA (mRNA), the most widely used indicator of cellular activity. Messenger RNA molecules are the genetic templates for proteins. In constructing proteins, the mRNA template is transcribed from genes and transported to the ribosomes - the cell's protein "factories" that are large complexes of protein and RNA.

"The basic problem is that, in the end, all the cell's functions are carried out by proteins, and proteins are encoded by messenger RNA," said Weissman. "But changing the mRNA level is only one of many mechanisms that the cell uses to alter levels of a protein." For example, the cell may produce and process proteins in numerous ways that ultimately affect the composition of the working molecule and its abundance. "The most abundant level of mRNA in a yeast cell is about a hundred molecules, and the lowest, non-zero level a cell can have is, by definition, one," he said. "But the functional level of protein in a cell can vary enormously, as much as a million-fold difference." Thus, measuring the levels of proteins in the cell is critical to understanding the functional properties of proteins.

Likewise, mapping the location of individual proteins in the cell is also critical to understanding protein function, said O'Shea. "A major goal of genomics and proteomics is to understand the function of each protein, and important clues can be gleaned from where each protein is within the cell," she said.

To measure the levels and locations of the thousands of individual yeast proteins, the researchers developed a method for tagging each protein. Without a common tag, measuring individual proteins is all but impossible, given that there are so many and each protein is biochemically unique. So, the researchers employed a widely used technique to introduce the DNA for a specific tag into each of the genes that specify each yeast protein. To create the targeted tags, they synthesized about 13,000 gene sequences that would target one of the tags to the end of each gene in the yeast genome. They used these specific tags to mark some 6,234 yeast gene segments, called "open reading frames."

As a result of the gene-tagging, the researchers created two different "libraries" of yeast cells. One library consisted of roughly 4,200 yeast "strains," each producing one tagged protein—tagged with a sequence that makes it easy to detect. This method permitted the researchers to use antibodies to quantify the level of protein in the cell. The other library, used for protein localization studies, consisted of strains where each protein is tagged with a sequence that produced a green fluorescent protein visible under a microscope.

"Fortunately, yeast cells have a peculiar property that enables us to target gene insertions to specific points on the chromosome," said O'Shea. "So, part of what's special about these experiments is that we could measure proteins

that are expressed from the chromosome under normal control. So, they are expressed at physiologically relevant levels and in relevant places in the cell.”

The researchers' protein-expression studies revealed the abundance of some 4,251 proteins in the yeast cell. “The experimental highlights for me are, first, that we can detect more than eighty percent of the yeast proteins in the cell—and that's a high fraction of the genome to be expressing at one time,” said O'Shea. “And equally interesting is that we can see a vast range of protein abundance, from fewer than fifty molecules per cell to more than one million.” In contrast, previous methods of detecting yeast proteins identified only the more abundant proteins, missing rare but critical proteins, such as those that switch on genes.

“We just didn't know how much of the proteome the cell needed during growth,” said Weissman. “We might have guessed that the cell had many genes in reserve for other purposes. So, the eighty percent expression level we detected was a bit of a surprise. Also, although we had hints that proteins existed at a wide range of abundances, until these measurements were made, there was no way to quantify that range.”

In addition to identifying many thousands of functional genes, the researchers also quantified the number of “spurious open reading frames,” which are DNA segments that appear to be genes, but which are not. Their findings, they said, agreed with previous studies of nonfunctional DNA segments by other researchers.

The researchers next plan to use their protein libraries to explore how protein levels change over time. Obtaining this kind of dynamic information will be critical to efforts to model the action of proteins as the cell grows and adapts to changing conditions. Such models will give biologists the scientific equivalent of a movie of the cell's machinery, rather than the snapshots available today, they said.

“Although it is interesting to know under standard lab conditions how much of each protein is present, what you really want to know in order to understand more about biology and biological processes, is how the amounts of the proteins change in response to different perturbations, like changes in the environment. Or if you make a mutation in the cell, how did the abundance of the proteins change,” said O'Shea.

The researchers' protein-localization studies revealed where some 4,156 proteins resided in the cell. According to Weissman, the localization studies were performed at two levels of specificity. “First, we examined the cells using microscopy techniques, and for many proteins that was enough. We could tell that many proteins—more than a thousand—existed in the nucleus or the cytoplasm.” The nucleus contains the cell's genetic material and the cytoplasm is the fluid-filled region outside the nucleus. “But for other

proteins, we saw a punctate pattern that only told us that the protein was concentrated in a specific place.”

To pinpoint the locations of these proteins more specifically, the researchers introduced genes for red fluorescent molecules that were known to localize to one or another specific cell structure. Such structures might include the mitochondria—the cell's power plant—or the golgi apparatus, which is a network of internal cell membranes. When the researchers saw both red and green fluorescence at a given point, they knew that the protein concentrated in that structure.

According to O'Shea, the results of the localization studies were gratifying, and are already beginning to have an impact. “We were surprised that we could see as many proteins as we did, and that the quality of the data was so good. Also, we were surprised that more than eighteen hundred proteins have at least a part of their localization in places other than the cytoplasm or the nucleus. So, from this study we've gained a lot of information about potential new functions of these proteins.”

O'Shea said that her future studies of localization, like those quantifying protein levels, will concentrate on dynamic changes in the cell. “In this study, we've only provided a static view of localization under one condition,” she said. “But protein localization is dynamic, in many cases, and I think that the big challenge now is to use this library of strains to study how protein localization changes in response to environmental conditions.”