

MAY 10, 2006

Putting More Protein Structures in the Research Pipeline

Biologists have long been thwarted in determining the three-dimensional structure of proteins that carry out their jobs only after intimately embracing other proteins. Although the structures of these complexes could reveal a bounty of new details about how proteins function, this information has been slow in coming because the work is difficult and time consuming.

Meanwhile, says Howard Hughes Medical Institute investigator David Eisenberg, the new methods of structural genomics have accelerated studies of individual proteins, but fail in many cases for proteins that are members of complexes.

"Our idea is that you could rescue from the trash bin of structural genomics many of the proteins that fall out of the pipeline."

— **David Eisenberg**

Eisenberg and his colleagues at the University of California, Los Angeles, decided the time was right to develop a new approach to tackling the structure of protein complexes. In an article published during the week of May 8, 2006, in the Early Online Edition of the *Proceedings of the National Academy of Sciences (PNAS)*, they describe the development of a genomic and structural-analytical approach that promises to rescue partner proteins from obscurity.

In the experiments presented in *PNAS*, Eisenberg, his graduate student Michael Strong, and others offer a proof-of-principle example of how this technique plays out in a real world application. They used their new approach to determine the structure of a protein complex unique to the tuberculosis bacterium. The fact that the protein complex, called PE/PPE, is only found in the bacterium *Mycobacterium tuberculosis* makes it a possible target for anti-tuberculosis drugs that would be more effective and specific than current antibiotics.

In the major NIH-sponsored structural genomics projects, a large fraction of proteins have been essentially lost from the research pipeline because either they are not expressed or they are expressed in an insoluble form, said

Eisenberg. A protein's solubility is a critical factor for further structural analysis by widely used techniques such as nuclear magnetic resonance spectroscopy and x-ray crystallography.

Eisenberg and his colleagues sought to develop a technique that would enable expression of both partners of a protein complex. The technique involves first performing computational genomic analysis to infer whether two proteins might be functionally linked. For example, if the genes for the proteins lie close together in the genome, they might be more likely to exist as partners in the cell.

Second, the researchers engineered the genes for the proteins so that they would be expressed together in the bacterium *E. coli*. They could then use co-purification techniques to determine whether the proteins form a paired complex. Finally, the researchers used standard structural analytical techniques to determine the structure of the purified protein pair.

Our idea is that you could rescue from the trash bin of structural genomics many of the proteins that fall out of the pipeline, and then restore them in complexes, said Eisenberg. That approach has two advantages. First, you learn the structure of the protein that you started out to study. But probably more important, because it's in a complex with other proteins — and you might know something about those other proteins — you begin to learn about its molecular biology because it's through protein-protein interactions that the cell operates.

In their proof-of-principle demonstration, the researchers determined the structure of a pair of members of two large families of proteins — called PE and PPE — found uniquely in the tuberculosis bacterium. While the proteins make up a significant fraction of the tuberculosis bacterial genome, their structures and functions remain unknown, said Eisenberg. Despite extensive efforts in Eisenberg's laboratory, neither the PE nor PPE proteins could be isolated.

Also, we thought this family might be a good family to target because there are no human homologs for them; therefore, drugs against those proteins might interfere with TB but would be less likely to produce side effects, said Eisenberg

Genomic analysis by the paper's first author, Michael Strong, revealed that the genes for members of the two protein families were often found near one another. Such proximity implied that the proteins would be produced together and would function together, said Eisenberg.

The researchers chose two particular members of the PE and PPE families that were very closely associated on the genome and had size and structural characteristics that seemed likely to make them easier to analyze. The researchers found that they could, indeed, express the two proteins together in *E. coli*. And when they purified the proteins together, they found clear evidence that the proteins existed only as a complex in the cell. Using the complexed proteins to produce crystals, the researchers performed x-ray

crystallographic analysis, developing the structure of the complex. This analysis revealed that the proteins' structures made them natural fits for one another.

The researchers also used computational analysis to compare the PE/PPE protein structure with proteins of known function, to deduce the possible role they might play in the cell. That analysis indicated that the complex had characteristics of proteins that function in a biological signaling pathway, perhaps in the cell membrane, said Eisenberg.

Eisenberg said that the findings with PE/PPE illustrate the value of the new approach to analyzing the structure of complexed proteins. To me, these findings say that protein partnering is a very sensible way to go in these structural genomics projects, he said. That is, you can identify those proteins that fail to be expressed individually and use computational genomic analysis to infer possible protein partners. And then on that basis, if there is a strong indication of one or more protein partners, you could co-express the protein with the others and determine whether they are, indeed, partners. The technique will help narrow the chasm between structural genomics and structural biology.

Although the functions of PE/PPE are still largely unknown, the new structural information offers hints of possible applications for treating tuberculosis. From such structures, one could try to design a compound that would interfere with the protein complex and keep it from forming, Eisenberg said. Interfering in such a way could kill the bacterium. And other critical proteins in the family might form the same kinds of complexes, so they may also be vulnerable to such compounds.

Eisenberg and his colleagues are now planning to apply their technique to a broader range of protein complexes. They are also exploring whether one partner protein might be used as a hook to draw another insoluble protein, such as a membrane protein, into solution for structural analysis.