

2nd Annual National Academies Keck *Futures Initiative* Conference

Designing Nanostructures at the Interface between Biomedical and Physical Systems

Arnold & Mabel Beckman Center, Irvine, California
November 18-21, 2004

Sequence a Single Molecule of Protein Focus Group Summary

Summary written by:

Maureen McDonough, Graduate Student, Science Writing Program, MIT

Focus group members:

- David Auston, President, Kavli Foundation
- Mark Hersam, Assistant Professor, Department of Materials Science and Engineering, Northwestern University
- Abraham Lee, Professor, Department of Biomedical Engineering, The University of California, Irvine
- Luke Lee, Professor, Department of Bioengineering, The University of California, Berkeley
- Randolph Lewis, Professor, Department of Molecular Biology, The University of Wyoming
- Hari Manoharan, Assistant Professor, Department of Physics, Stanford University
- Maureen McDonough, Graduate Student, Science Writing Program, MIT
- Thomas Perkins, Associate JILA Fellow, JILA, National Institute of Standards and Technology and The University of Colorado at Boulder
- Jon Pratt, Manufacturing Metrology Division, National Institute of Standards and Technology
- Alan Russell, Director, McGowan Institute for Regenerative Medicine, University of Pittsburgh
- David Tennenhouse, Vice President, Corporate Technology Group, Intel Corporation

Summary:

The protein sequencing group considered themselves lucky. With a clearly defined problem in hand, several members came to the first session with ideas about what the solution should look like. Jotted down on hotel letterhead the night before, it was clear that several of these eleven men wanted their solution on the fast track to the final presentation. The thought of finishing early and taking off to Disneyland was considered, but was taken off the table when everyone began to realize that not everyone had the same answer to the problem.

The group's problem was to figure out how to sequence a single protein molecule. The sequencing techniques currently used by researchers require a large and highly concentrated sample of a protein. However, many proteins exist naturally in extremely small quantities: an individual cell may only have one or two copies of specific hormones and transcription factors. The ability to sequence these proteins would help in determining their structure; and, by combining sequence and structure information, large amounts of a specific protein could be produced and used in therapies. Such techniques could also be used diagnostically by identifying specific proteins associated with conditions or diseases.

There was little debate about the focus group's goal or its importance, but deciding what was the best line of attack proved challenging. There were many strong personalities; as a result, no one person was able to force his vision upon his colleagues. Some group members were in favor of "visualization" and believed that if a protein could be linearized and attached to a solid surface without any contamination then the sequence could be read using an atomic force microscopy. Other members were in favor of the "flow channel" method in which a protein would be linearized and passed through a channel that would detect the sequence via nano-array. Someone else continued to stress the importance of using information available from the sequence of the human

genome, by checking the determined amino acid sequence against known DNA sequences.

The differences in opinion regarding the ideal solution led to a tendency for individuals to interject with statements or questions that would pull the discussion toward the idea in which they were most interested. Even with this tug-of-war, most of the group's discussions were very productive and focused on specific aspects of the problem. Eventually an agreement was reached to focus on the solution that was showcased in the group's final presentation. It was not a coincidence that this solution involved input from most of the group.

The first decision to be made was whether the protein should remain intact, throughout the sequencing process, or if each amino acid should be systematically cleaved and detected. One of the benefits of keeping the sequence intact is that a single protein molecule could be sequenced many times. The group members in favor of visualization and the flow channel solutions cited repeatability as a huge benefit to their approaches. However, it was estimated that the visualization technique would take a trained technician an entire day to sequence a single protein, and a more efficient method was desired. The flow channel solution was also ruled out because most of the group was convinced that the forces exerted on the linearized protein as it passed through the channel would break it apart. So the group was forced to deal with the fact that the protein would need to be chopped up. There would be only one opportunity to read the sequence and then "game over."

The group decided that the chopping reactions currently used in sequencing could be used in single molecule sequencing as well. Using specific chemical reactions, individual amino acids could be cleaved from the amine end of the protein one at a time. In the first step of the group's design, the protein would be bound to the sample chamber at the carboxyl terminus. There was concern about losing the single protein molecule in this step, so it was suggested that a fluorescent probe could bind to the protein. Once detected, the stepwise reaction would cleave off a single amino acid to be identified.

The free amino acid would then be washed down into the first detection chamber. Here the amino acid would be temporarily bound to a silver substrate, and a laser would be used to generate a surface enhanced Raman spectra. SERS detection provides some information about the structure of a molecule, and may be able to determine the specific amino acid, but at a minimum could be used to ensure that an amino acid was released during the cleaving reaction, which only goes to 99.8% completion.

The amino acid would then be washed down into the second detection chamber called the riboswitch chamber. A riboswitch is a sequence of RNA that cuts itself in half when a specific molecule binds to it. Two naturally occurring riboswitches have been discovered that are tripped by glycine and lysine, respectively; and the group suspected that switches specific to the remaining 18 amino acids could be engineered. The riboswitches that are specific to each of the 20 amino acids would be attached along a wall of this detection chamber. Like balloons on strings, attached to each variety of riboswitch would be a specific colored quantum-dot. Also called Qdots, these semiconductor nanocrystals light up in a variety of colors.

All of the lysine switches, for example, could be red; and all of the glycine switches could be green. When the amino acid enters the chamber it would bind to its specific riboswitch, which would then cleave itself; and a specific colored Qdot would be released. A sensor would detect the color, and the amino acid could be identified and then compared to the Raman spectra results. The amino acid and the Qdot are then washed out of the chamber and another amino acid could be cleaved in the sample chamber. The determined sequence would then be compared to known sequences in a database.

As creative and colorful as this idea is, the group identified several places where the mechanism could break down. One concern that deeply bothered the group was how to ensure that the amino acid did not get stuck to a wall on its way through the detection chamber. It was suggested

that a solution with a high salt concentration could be used to wash off a stuck amino acid, but that could raise stability problems with the surface/riboswitch/Qdot complexes.

Research challenges were identified at each of the steps in the group's solution. The sample chamber needs to be scaled down to a single molecule. The SERS detection chamber needs an optimized surface substrate, and Raman signature spectra for each of the 20 amino acids needs to be determined. The riboswitch chamber needs switches and quantum dots specific to all 20 amino acids and a way to attach the switches to the chamber wall and the dots to the switches.

The second solution presented was the rejection of the central dogma. More dramatic and perhaps less practical of a solution, some members of the group hope to achieve "reverse translation." The argument presented was that though an enzyme that could achieve reverse translation has not been identified, it may exist somewhere in nature. After all, no one believed that reverse transcription was possible until it was discovered that nature had found a way. Even if the enzyme could not be found in nature right now, perhaps it could be created in a laboratory. In order to work, the enzyme would need to be able to use tRNA to identify each amino acid in order and ligate the RNA codons to generate the mRNA.

Another solution involved the riboswitch model. If DNA could be released instead of a quantum dot, then each piece of DNA representing an amino acid could be ligated to the previous piece, creating a sequence of DNA that corresponded to the amino acid sequence. The standard procedures for DNA sequencing could then be used and, in effect, reverse translation achieved.

What made this focus group unique was the specificity of the problem. Because there was no real debate about what the problem was, there was time to address several possible solutions and their individual challenges. The design and the discussions were about details. And though they did not get a chance to go to Disneyland, I think everyone was happy with the focus group's findings.