Use of Protein Barcodes and Next-Generation Protein Sequencing™ for High-Throughput Mapping of Protein Conformations and Their Phenotypic Effects

Eugene Serebryany Ph.D.
Assistant Professor, Department of Physiology and Biophysics
Stony Brook University

"There's a vast conformational space of proteins that remains to be explored due to a lack of high-throughput experimental methods. We're advancing a novel technique that combines protein barcodes and next-generation, single-molecule protein sequencing to enable systematic, high-throughput mapping of non-native conformations that can't be crystallized, and then link them to phenotypic effects. The combination of single-molecule sequencing and protein barcodes has the potential to dramatically expand this field of study."

Proteins can adopt a range of distinct conformations, some of which drive the pathology of many diseases. Protein misfolding and the resulting aggregation has been implicated in neurodegenerative conditions such as Alzheimer's, Parkinson's, and Huntington's disease, as well as non-neurological diseases including cataracts, some forms of atherosclerosis, short-chain amyloidosis,[1] and cancer.[2]

Misfolded proteins can exist as soluble monomers, small aggregates, or as large insoluble inclusion bodies. While the underlying biophysical mechanisms leading to cytotoxicity are not fully defined, possible explanations include stress on cellular quality control and protein degradation systems, disruption of cell membranes, and aberrant protein-protein interactions.[3] A better understanding of the toxic effects and clearance of misfolded proteins may reveal novel strategies for the prevention and treatment of many diseases. Elucidating non-native, transient, and disordered protein conformations has proven challenging with current experimental techniques, as they cannot be crystallized, a method that is commonly used to define the structure of proteins.

Eugene Serebryany, Ph.D., and his team at Stony Brook University in New York have developed an innovative approach to systematically discover, stabilize/trap, and purify native and non-native protein conformations generated *in vitro* or in *vivo*, and directly link them to molecular, organismal, or evolutionary phenotypes.

We spoke with Dr. Serebryany about the limitations of conventional methods to study misfolded proteins and how next-generation protein sequencing (NGPS) on the Quantum-Si Platinum® instrument and protein barcodes are poised to dramatically expand this field of study.

Q: What is the focus of your research?

ES: My main interest is the biophysics of protein misfolding in vivo and mechanisms of misfolding-associated diseases. I'm also interested in the discovery and pharmacological targeting of physiologically relevant non-native protein conformations, and protein engineering in currently inaccessible environments. My lab is addressing fundamental questions about how we can explore the conformational space of proteins, how we can trap them in different conformational states, and how to map what those conformational states do.

To enable and facilitate these areas of research, we develop new experimental methods such as high-throughput disulfide scanning (HTDS) of protein conformations. This technique allows us to map conformational landscapes to phenotypic landscapes *in vivo*.

Q: What was the inspiration to develop HTDS?

ES: Let's start with a bit about disulfide bonds. Disulfide bonds between cysteine residues in proteins create conformational constraints and help maintain the stability of proteins. Protein disulfide isomerase is an enzyme found in the endoplasmic reticulum (ER) in cells and the periplasm of bacteria that catalyzes the formation and breakage of disulfide bonds as proteins fold. This process allows proteins to assume the correct, fully folded state. The same enzyme also scavenges misfolded proteins in the ER.

Disulfide scanning mutagenesis is an established technique for determining which double-cysteine variants of a protein form intramolecular disulfide bonds. The method is typically used for testing structural models of proteins. It has limited throughput, which is problematic as the number of double-cysteine variants increases significantly with longer length polypeptides. Scanning an entire protein has never been practical which is why we developed HTDS, a qualitatively new capability that enables mapping conformational landscapes onto phenotypic landscapes.

Q. How does HTDS work?

ES: We adapted disulfide mapping, which is a process where you put a pair of cysteine residues somewhere in the structure of a protein. If the residues are close together, under some conditions they will form a disulfide bond; if they are far apart, they won't.

To reveal which disulfides trap which chromatographically resolvable conformers, we devised a deep sequencing method for double-cysteine variant libraries of proteins that precisely and simultaneously locates both Cys residues within each polypeptide. We can essentially unpack the 3D structural space of a protein into subsets of one-dimensional sequence space via disulfide cross-links and map whether or not a given pair of cysteine residue is able to covalently link to each other.

We published this method in *Molecular Cell* in 2023 and as a proof of concept, we applied the technique to HdeA, an intrinsically disordered protein chaperone found in *E. coli* periplasm.[4] HdeA inhibits aggregation of periplasmic proteins as *E. coli* passes

through stomach acid on its way to the gut. An intramolecular disulfide links the only two native cysteine residues; reduction of this bond disorders the structure of the protein and inhibits chaperone activity *in vitro*. Our study demonstrated the feasibility of disulfide scanning of an entire protein in one experiment.

It's a remarkably successful approach because it allows us to not only look at the biophysical properties of protein conformations in vitro, but also the phenotypic consequences *in vivo*. We can basically transform these double cysteine scanning libraries into the host organism, in this case *E. coli*, or different cell types, and determine, at the phenotypic level, which cells have greater fitness, which have lower fitness, which have different phenotypes that can be sorted, and which disulfide corresponds to those phenotypes.

At the same time, *in vitro*, we can assess the changes in the biophysical properties of the protein molecule that arise from the formation of specific disulfides that trap specific types of conformations.

Q: What impact will the use of protein barcodes have on this method?

ES: In our initial proof of concept, we had to rely on mass spectrometry for the protein sequencing, which is possible, but very challenging. Mass spec is very good for point mutations, but with disulfide scanning, you need to make two mutations at a time in the same molecule, and you need to know the location and the sequence of both cysteines in the same polypeptide chain at the same time.

We had to adapt a rather clever type of chemistry that allowed us to cleave proteins specifically at the sites where the cysteines were, and then mass spec the intervening peptides to map the termini, and therefore where the cysteines used to be. While that works to a certain extent, it has limitations. The cysteine residues can't be not too far apart in the sequence because mass spec cannot handle very long peptides. Also, there can't be any intervening cysteines between the two that you've introduced, because that would lead to a cut between the two introduced cysteines, leading to a complete loss of all the structural information.

These limitations can be overcome with the use of protein barcodes, which also would allow application of this technique to larger, more complex proteins and those with native cysteine residues. We could create a library of peptide barcodes attached as fusions to one of the termini. At the end of the experiment, we simply cut off the barcode and sequence it.

Q: What barcoding strategy would be preferable for HTDS?

ES: The ideal barcode is as short as possible with all the barcodes in the library having similar amino acid composition and chemical properties – so, perhaps just the same combination of residues, or a few chemically similar combinations, with the residues rearranged in various orders. This enables a large library of barcodes for increased throughput while minimizing the likelihood of differential effects on protein structure and folding, and therefore minimizing the chances of getting artifacts in the experiment. The sequence of the barcode should be as short as possible, maybe four, five, or six residues, depending on the complexity of the library that is needed, of course.

Barcodes with the same amino acid composition are difficult to sequence using mass spec because they would yield the same single precursor ion. Using barcodes that are sufficiently different would increase the chemical heterogeneity of the barcode library; barcodes with different chemical properties (such as hydrophobicity) would potentially interfere with the protein structure in different ways. Moreover, mass spec typically requires peptides longer than six residues for confident sequence determination.

Single molecule peptide sequencing is therefore preferable to mass spec for analysis of libraries of short, chemically similar peptide barcodes. This is what really piqued my interest in Quantum-Si, and we are now leveraging this technology for HTDS. There's a lot of innovation here that can further advance our novel structural technique. Exploring the vast conformational space of proteins via peptide barcoding required a new sequencing approach, and we turned to the Quantum-Si instrument to find it.

Q: How do you envision HTDS using next-generation protein sequencing and protein barcodes being leveraged in drug discovery?

ES: I think the potential is quite expansive. This technique enables many types of biophysical measurements such as thermodynamic stability, binding affinity to a target, or aggregation propensity, at much higher throughput than conventional methods. Because HTDS stabilizes diverse conformations, it may prove valuable in applications such as vaccine design and drug screening. An example we noted in our *Molecular Cell* paper was related to the respiratory syncytial virus (RSV) vaccine. Stabilizing the wrong conformation of the antigen in the vaccine led to tragic consequences[5]; disulfide engineering was subsequently used to stabilize the correct conformation.[6] The same approach was

applied to the SARS-CoV-2 spike protein, but less than 100 disulfides were screened, chosen by some combination of structural intuition and computational modeling.[7] HTDS could allow screening of many thousands of disulfides even before an atomistic structure is available and may reveal disulfides with unexpected useful allosteric effects.

HTDS could also advance treatment of neurodegenerative and other diseases that result from protein misfolding. Stabilization of misfolded aggregation precursors via disulfide cross-linking could be used as protein vaccines. Another interesting application is the identification of disulfides that are able to trap non-native cytotoxic conformations. This could enable screening for drugs that select that conformation, offering the potential to create a new class of antibiotics that work by induced misfolding of target proteins.

REFERENCES

1. Reynaud, E. (2010) Protein Misfolding and Degenerative Diseases. Nature Education 3(9):28
2. Li J, et al. (2022) p53 amyloid aggregation in cancer: function, mechanism, and therapy. Exp Hematol Oncol 11, 66. https://ehoonline.biomedcentral.com/articles/10.1186/s40164-022-00317-7
3. Folger, A., and Wang, Y.C. (2021). The cytotoxicity and clearance of mutant huntingtin and other misfolded proteins. Cells 10, 2835. https://pubmed.ncbi.nlm.nih.gov/34831058
4. Serebryany, E, et al. Systematic conformation-to-phenotype mapping via limited deep sequencing of proteins. Molecular Cell 83(11): 1936-1952. https://www.researchgate.net/publication/360144922_Systematic_Conformation-to-Phenotype_Mapping_via_Limited_Deep-Sequencing_of_Proteins
5. Killikelly AM, et al. (2016). Pre-fusion F is absent on the surface of formalin-inactivated respiratory syncytial virus. Sci. Rep. 6, 34108. https://doi.org/10.1038/srep34108
6. McLellan JS, et al. (2013). Structure-based design of a fusion glycoprotein vaccine for respiratory syncytial virus. Science 342, 592–598. https://pubmed.ncbi.nlm.nih.gov/24179220/
7. Hsieh, CL, et al. (2020). Structure-based design of prefusion-stabilized SARS-CoV-2 spikes. Science 369, 1501–1505. https://doi.org/10.1126/science.abd0826