Nanopore-based Protein Identification

Lead Group: Francesco De Angelis, Plasmon Nanotechnologies, IIT
Partner Group(s): Nick Goldman, EMBL-EBI

Rationale & Hypothesis: The ability to identify the human proteome at the level of single individual can significantly influence our comprehension of Life and our ability to prevent, diagnose, and treat many pathologies. Conventional approaches for detection and characterization of proteins (such as mass spectroscopy, X-ray crystallography, NMR and ELISA) are very difficult to apply to single individuals because they are cost- and time-consuming. This hindered their use for the development of personalized approaches. To this aim, it is necessary to identify epigenetic, post-transcriptional, and post-translational modifications that are specific to single individuals. It is also essential to develop the ability to identify low abundance proteins in a complex matrix, a current challenge in protein biomarker discovery. In these contexts, nanopore technologies based on electrical readouts are emerging as promising methods [1]. However, among different limitations these methods show limited sensitivity (far from single molecule level) and poor discrimination power (only a few types of amino acids can be recognized by labelling). Notably, electrical readout methods intrinsically suffer from molecular translocation occurring too quickly for reliable measurements to be taken: typically on a time scale of ns–μs where signal-to-noise ratio is a significant issue. However, recent tremendous advances in optical systems such as ultrafast lasers and ultrafast/ultrasensitive cameras make this time scale now largely accessible even at single photon level. In this regard, the exploitation of Raman spectroscopy for protein investigation may represent a key enabling technology, thanks to its ability to discriminate the amino acids and nucleotides through their unique spectral fingerprints [2,3] (with no labelling). In particular, when combined with plasmonics and nanopores, Raman reaches the extreme sensitivity and high speed needed for single amino acid detection [2,3], thus potentially enabling single protein identification and sequencing.

Aims: The objective of this research is to develop an ultrafast Raman platform able to record the amino acid sequence along a protein chain on a time scale of 100 ns, thus enabling the identification and sequencing of human proteins with unprecedented performance not reachable with existing approaches. To this aim, we will combine innovative plasmonic nanopores with next generation ultrafast time-resolved single-photon detectors. State-of-the-art machine learning techniques [4] in combination with bioinformatics approaches [5] will be developed to analyze the high-throughput data output with multiple specific aims/steps: 1) reconstruct the Raman spectra from few photons; 2) use the Raman spectra to identify the sequence of amino acids translocating through the pore; and 3) identify the protein/peptide by the sequence of amino acids previously identified.

Significance & Impact: The development of a simple, fast, sensitive (few molecules) and low-cost technique for protein identification and sequencing would open a new paradigm in proteomics and could bring the early diagnosis assays and personalized treatments envisioned for future medicine. Notably, the development of ultrasensitive Raman technologies able to analyze liquid samples in flow-through schemes and with high throughput rates will pave the way to a new generation of detection devices suitable for many different purposes, not only the future protein market but also food analyses, quality control and safety assessment of market compounds, water pollution, and many others**.**

Integration of Expertise of Partners: The Plasmon Nanotechnologies group at IIT provides state of the art techniques in Raman measurements at single molecule level and flow-through schemes. The Goldman group at EMBL-European Bioinformatics Institute provides advanced DNA and protein sequence analyses which are an essential skill for the future generation of PIs working on single molecule spectroscopy and biology, including novel algorithms for nanopore sequencing.

Person Specification:
We expect the successful candidate to be based primarily at the Plasmon Nanotechnologies group, IIT, Genova, Italy. Some time would be spent working with the Goldman Group at the EMBL-European Bioinformatics Institute, Cambridge, UK.

Notes:
[1] Restrepo-Pérez et al., 2018, *Nature Nanotechnology* **13**:786–796.
[2] Huang et al., 2019, *Nature Communications* **10**:5321.
[3] Huang et al., 2020, *Angewandte Chemie Int. Ed.* doi.org/10.1002/anie.202000489.

[4] See, e.g., Wick et al., 2019, *Genome Biology* **20**:129 and references therein for discussion of analogous methods in DNA sequence analysis.
[5] For example, adapted from ideas familiar in mass spectrometry, e.g. Wang & Wilson, 2013, *BMC Bioinformatics* **14**:S24.