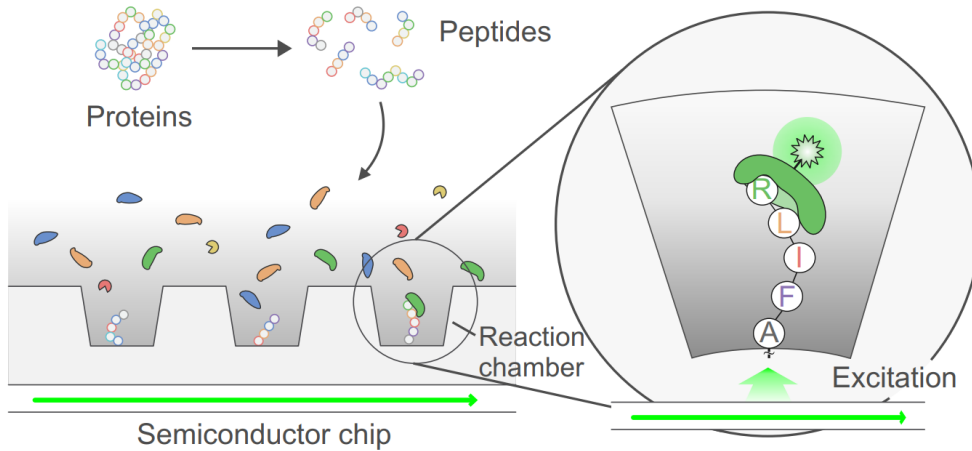


# Real-time dynamic single-molecule protein sequencing on the Quantum-Si platform

# Real-time single-molecule protein sequencing



## Library Prep

- Protein digestion
- Functionalization at C-terminal lysines

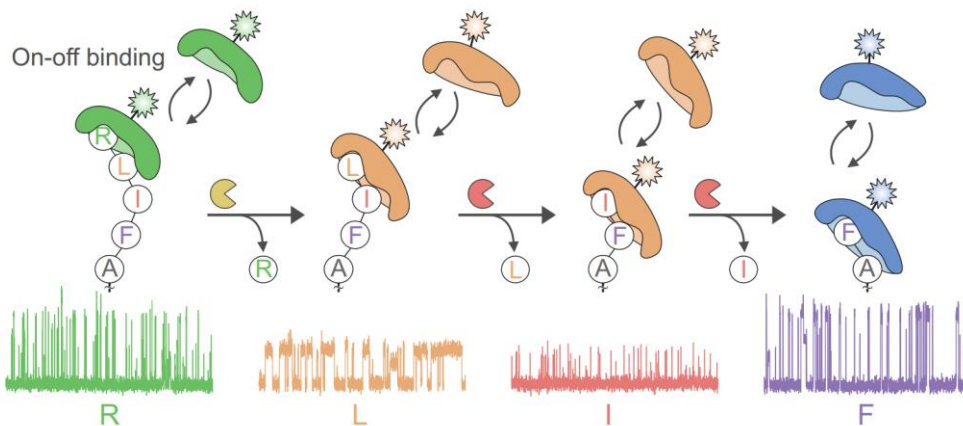
## Loading

- SV-mediated surface attachment at bottom of nanowells
- Poisson distributed
- Excitation light delivered to loaded peptide complex from nearby waveguide

Recognizers



Aminopeptidases



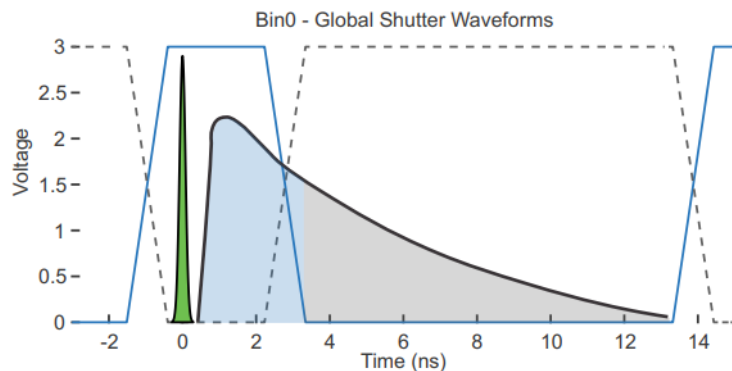
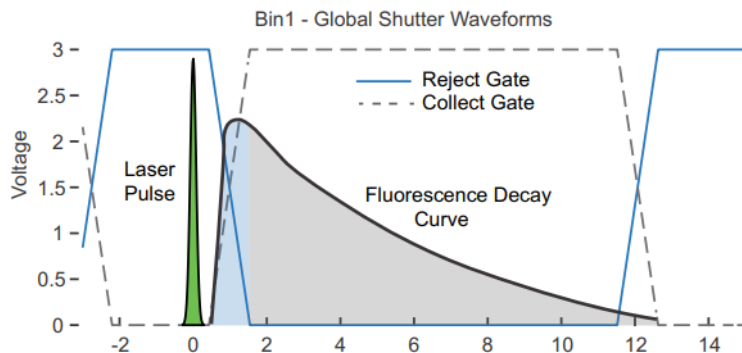
## Real-time single-molecule protein sequencing

### Recognition

- N-terminal recognizers labeled with different fluorescent dyes
- Recognizers bind one or more N-terminal amino acids (NAAs)
- 10s-100s of pulsing events per amino acid

### Cleavage

- Aminopeptidases perform stepwise NAA cleavage
- Cleavage events stochastic at the single-trace level



## A time-domain sensitive chip

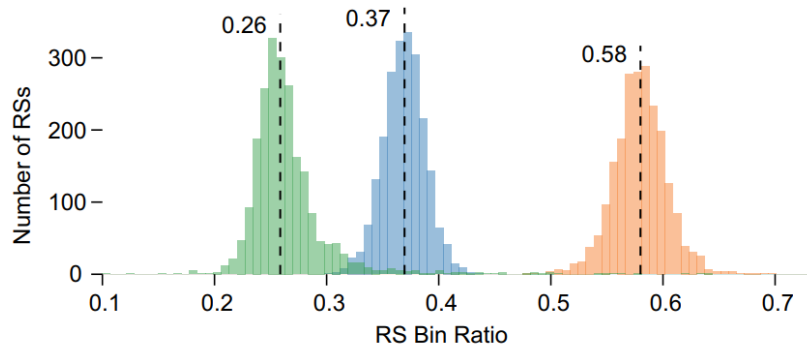
### Laser rejection and Fluorescence lifetime collection

- Integrated 532 nm pulsed laser
- Chip cycles between laser rejection and collection windows. ns scale.
- Early and late start of collection captures different portions of the fluorescence decay curve in alternate frames. ms scale.
- Fluorescent dyes with different lifetimes distinguishable by "bin ratio" and intensity

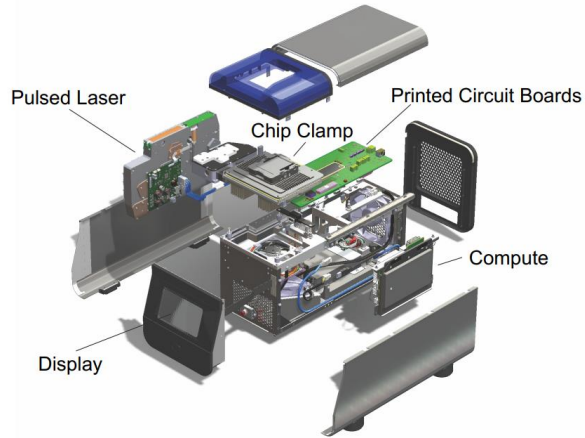
## A time-domain sensitive chip

### Laser rejection and Fluorescence lifetime collection

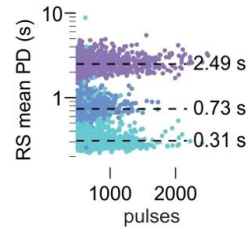
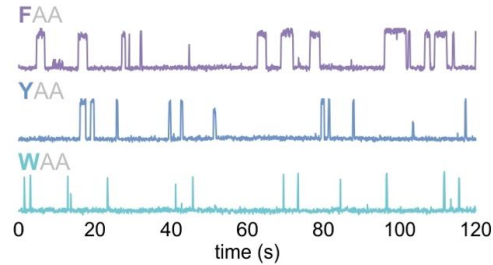
- Integrated 532 nm pulsed laser
- Chip cycles between laser rejection (Bin 0) and collection windows (Bin 1)
- Early and late start of collection captures different portions of the fluorescence decay curve in alternate frames
- Fluorescent dyes with different lifetimes distinguishable by "bin ratio" and intensity



## The Platinum instrument



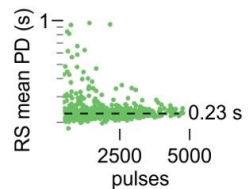
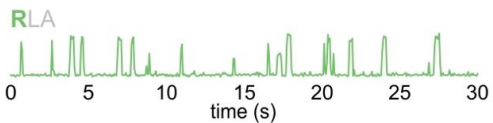
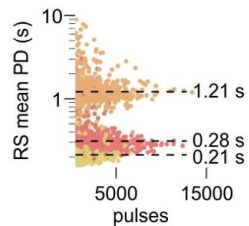
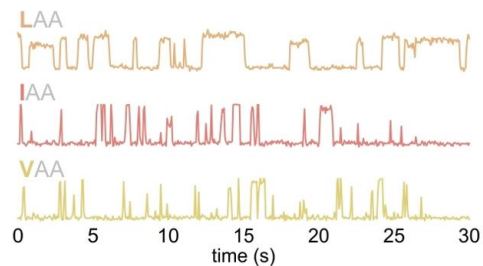
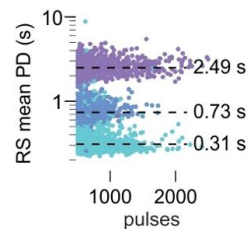
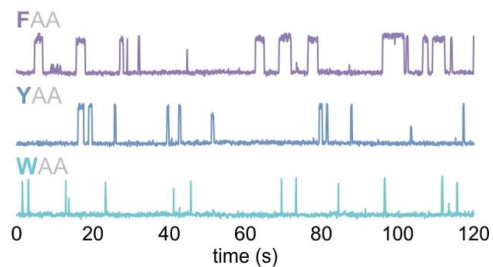
- Elimination of filtering and dye discrimination by wavelength enables reduced size, cost, and complexity
- Laser module and compute for signal processing integrated in instrument
- Production chips have 2M active wells, scalable to 10s of millions in first product line



## NAA recognition

### N-degron pathway proteins provide scaffolds for recognizer development

- First recognizer = PS610 derived *A. tumefaciens* ClpS2
- Visible pulsing on-chip for N-terminal F, Y, W
- Decreasing affinity reflected in average pulse duration for FAA, YAA, WAA peptides
- Average pulse durations (PDs)  $\sim 0.3$  to  $2.5$  s

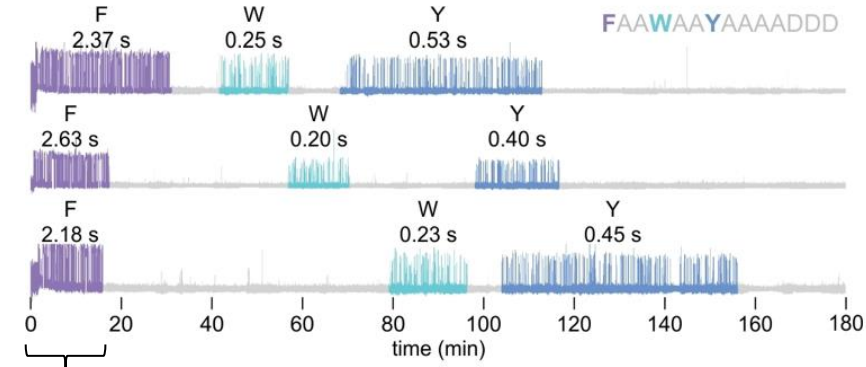


## NAA recognition

### N-degron pathway proteins provide scaffolds for recognizer development

- PS961 – Derived from a novel group of ClpS proteins from Planctomycetes that binds N-terminal L, I, and V
- PS691 – A UBR-box protein from the yeast *Kluyveromyces lactis* recognizes R visibly

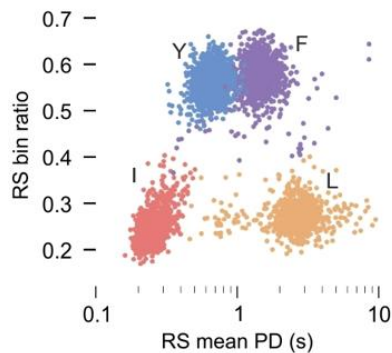
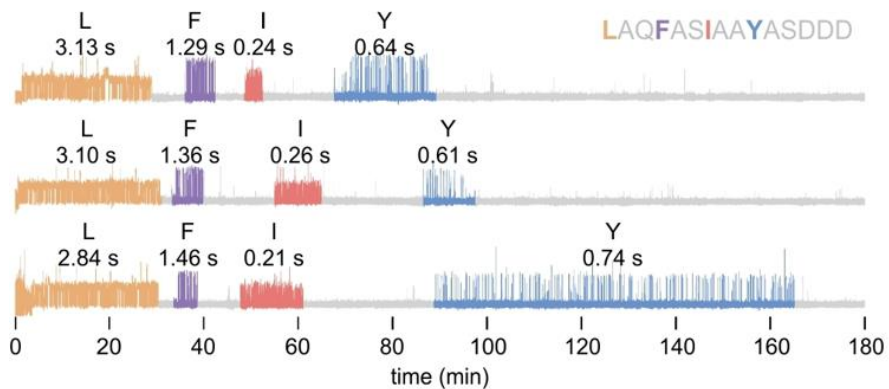
## Dynamic sequencing



recognition segment

### First demonstration of ordered recognition and cleavage

- Chip loaded with FAAWAAYAA peptide and PS610 (FYW) added
- A TET aminopeptidase from *Pyrococcus horikoshii* added at 15 min
- Traces display distinct recognition segments (RSs) for F, W, and Y, in the correct order

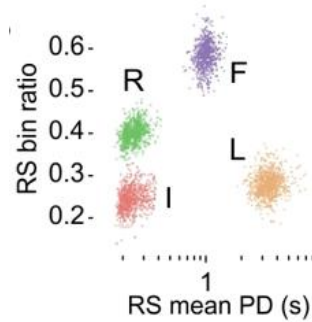
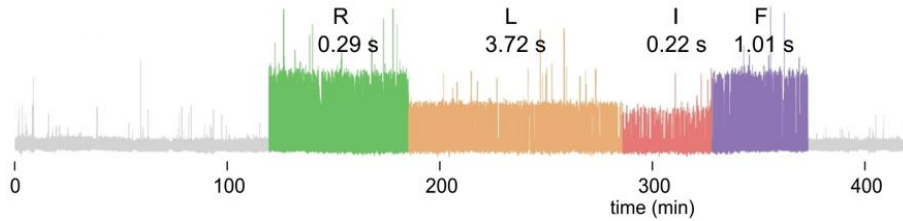


## Dynamic sequencing

### Demonstration with two recognizers

- PS610 and PS961 labeled with dyes that are distinguishable by lifetime
- Traces display distinct recognition segments (RSs) for L, F, I, and Y, in the correct order
- Bin ratio distinguishes recognizers, kinetics distinguishes NAAs

DQQRLIFAG



## Dynamic sequencing

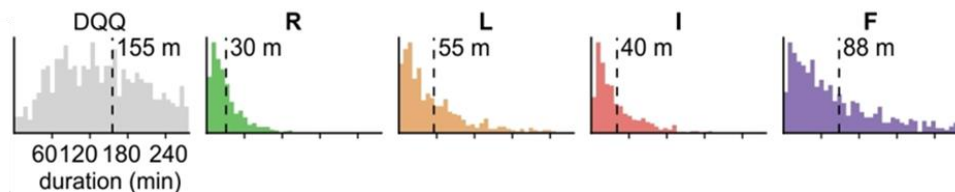
### Ubiquitin peptide illustrates the kinetic principles of the sequencing assay

- DQQRLIFAG = a segment of human Ubiquitin
- 3 Recognizers and 2 TET aminopeptidases



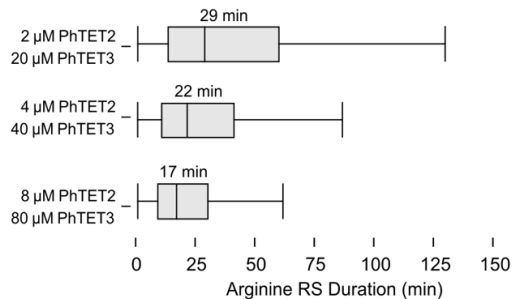
### N-terminal cleavage events are very fast

- <1 to a few seconds between RSs



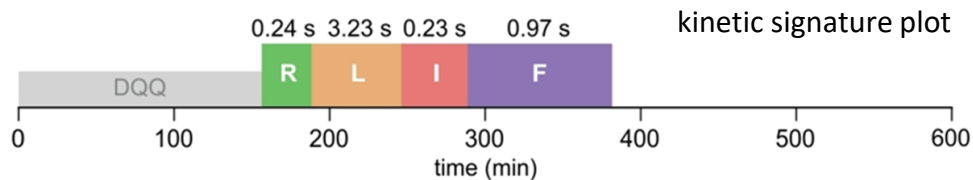
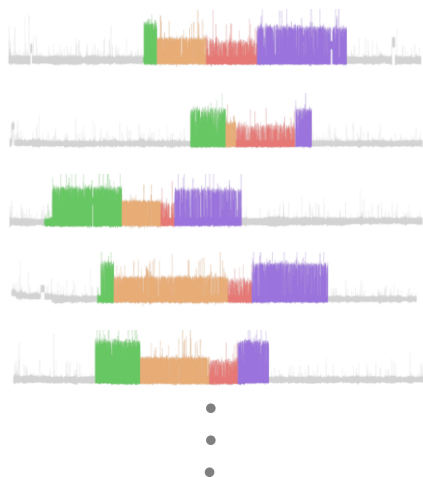
### Ensembles of traces reveal characteristic distributions of RS duration

- Approximate single-exponential decay statistics



### Aminopeptidase concentration controls RS duration

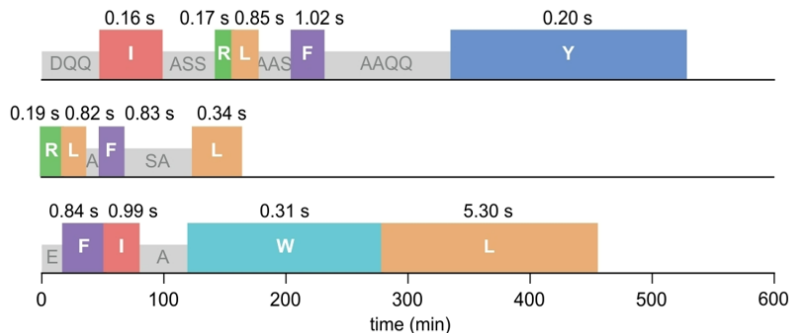
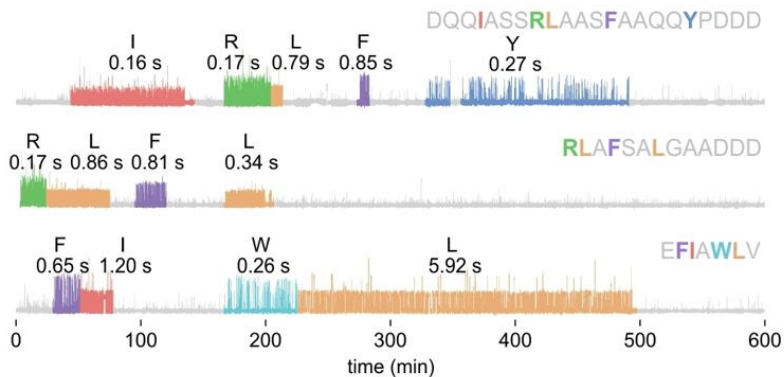
- Target average 10-40 min between cleavage events provides enough time for pulsing data collection and avoiding missed RSs



## Dynamic sequencing

**The kinetic signature plot summarizes the average sequencing behavior of an ensemble of single peptide molecules**

- Highly characteristic for different peptides
- Sensitive to sequence composition and PTMs
- Predictable for any peptide

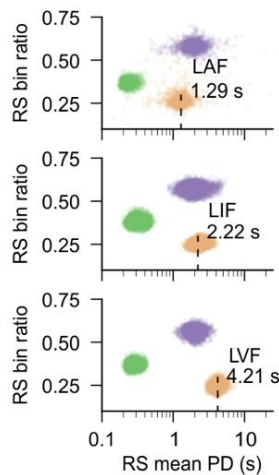
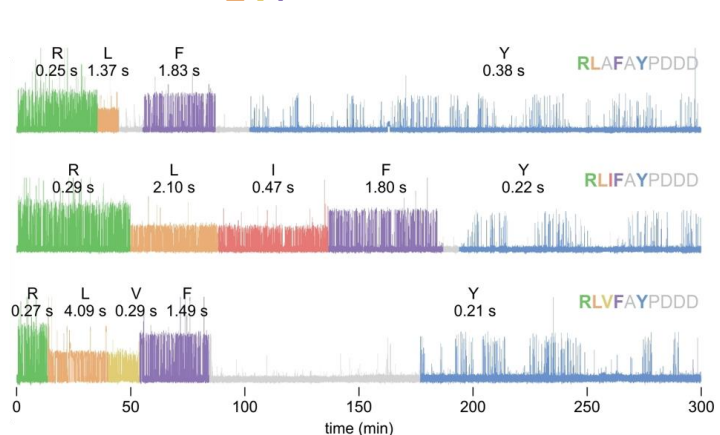


## Dynamic sequencing

### Sequencing diverse synthetic peptides

- The assay works with peptides across the wide range of sequence composition, physiochemical properties, and lengths expected in a proteome digest

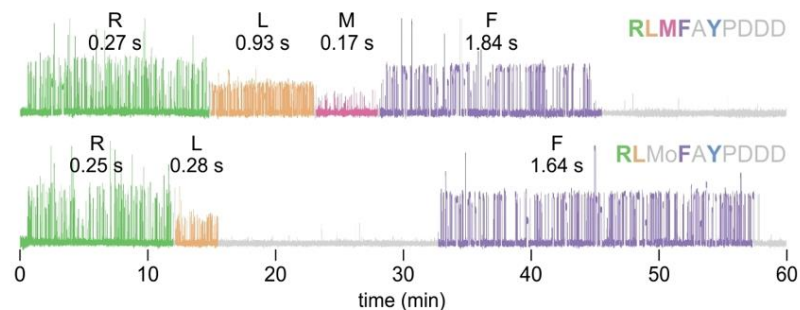
↓  
LAF  
LIF  
LVF



## Sensitivity to substitutions and PTMs

### Detecting single amino acid changes

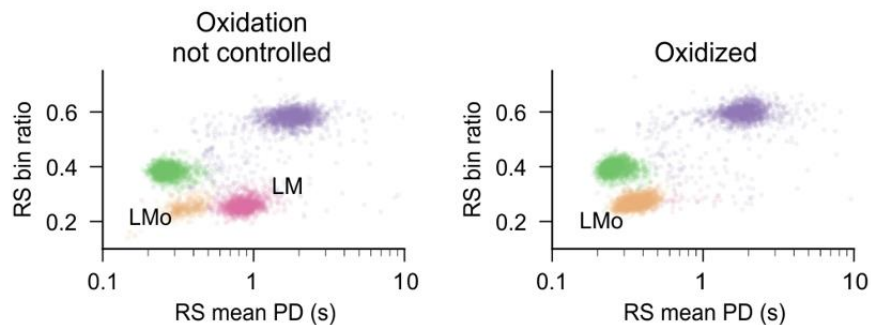
- Recognizers physically contact/sense residues downstream of bound NAA (all 20 types of AAs)
- Downstream sequence, predominantly P2 and P3, affects average PD
- Influence is encoded in the peptide's kinetic signature
- Positions can be sensed multiple times by different recognizers as sequencing progresses



## Sensitivity to substitutions and PTMs

### Detection of methionine oxidation

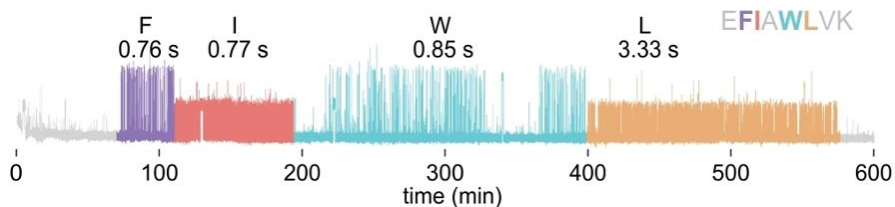
- Methionine oxidation (a single atom change) produces a highly visible change in sequencing output
- PD of preceding L decreases
- Eliminates recognition of methionine by PS961
- Provides a blueprint for general PTM detection
- We are able to detect other PTMs such as phosphorylation and methylation



## Ubiquitin digest:



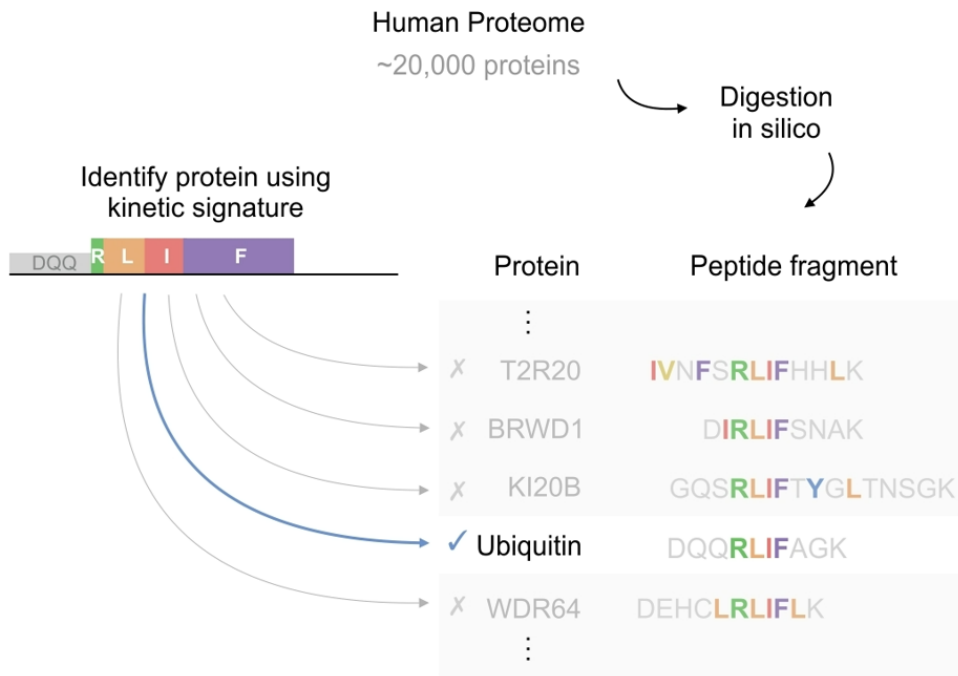
## Glucagon-like peptide 1 digest:



## Sequencing real protein libraries

### Examples: Human Ubiquitin and Glucagon-like peptide 1

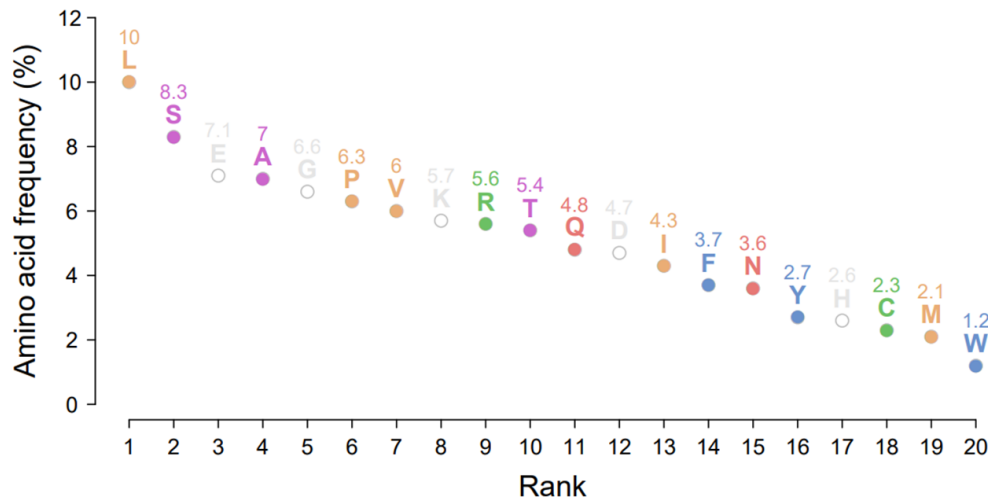
- End-to-end runs with recombinant Ubiquitin and GLP1 produce the expected peptides on-chip
- Kinetic patterns match those obtained from synthetic peptides



## Sequencing real protein libraries

### Mapping sequencing output to the proteome

- The information-rich sequencing output allows precise mapping of even short reads to their proteins of origin



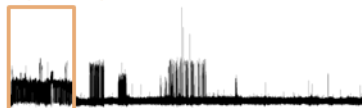
## Advances in proteome coverage

### Development of new and improved recognizers

- New recognizers developed through our protein engineering and directed evolution program
- Direct N-terminal recognition of 15 types of amino acids demonstrated
- Path to >70% direct recognition of the proteome sequence space, and interrogation of >90%
- Enables unique identification of 90% of proteins in the human proteome.



L (10.0%)



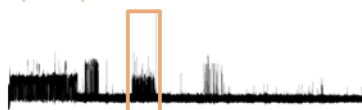
P (6.3%)



V (6.0%)



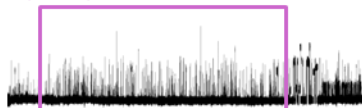
I (4.3%)



M (2.1%)



S (8.3%)



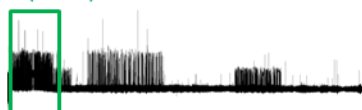
A (7.0%)



T (5.4%)



R (5.6%)



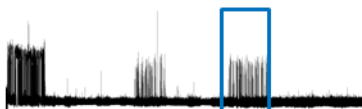
C (2.3%)



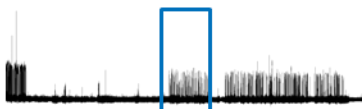
F (3.7%)



Y (2.7%)



W (1.2%)



Q (4.8%)



N (3.6%)



## Special thanks

### **QSI R&D Team**

Valentin Abramzon  
Omer Ad  
Pat Adcock  
Faisal R. Ahmad  
Gün Alppay  
James A. Ball  
James Beach  
Dominique Belhachemi  
Anthony Bellofiore  
Michael Bellos  
Juan Felipe Beltrán  
Andrew Betts  
Mohammad Wadud Bhuiya  
Kristin Blacklock  
Robert Boer  
David Boisvert  
Norman D. Brault  
Aaron Buxbaum  
Steve Caprio  
Changhoon Choi  
Thomas D. Christian  
Robert Clancy  
Joseph Clark  
Thomas Connolly  
Kathren Fink Croce  
Richard Cullen

Mel Davey  
Jack Davidson  
Mohamed M. Elshenawy  
Michael Ferrigno  
Daniel Frier  
Saketh Gudipati  
Stephanie Hamill  
Zhaoyu He  
Sharath Hosali  
Haidong Huang  
Le Huang  
Ali Kabiri  
Gennadiy Kriger  
Brittany Lathrop  
An Li  
Peter Lim  
Stephen Liu  
Feixiang Luo  
Caixia Lv  
Xiaoxiao Ma  
Evan McCormack  
Michael J. Meyer  
Michele Millham  
Roger Nani  
Manjula Pandey  
John Parillo  
Gayatri Patel

Douglas H. Pike  
Kyle Preston  
Kyle Rearick  
Todd Rearick  
Gerard Schmid  
Jonathan Schultz  
Xinghua Shi  
Badri Singh  
Nikita Srivastava  
Shannon F. Stewman  
T.R. Thurston  
Philip Trioli  
Jennifer Tullman  
Xin Wang  
Yen-Chih Wang  
Eric A. G. Webster  
Zhizhuo Zhang  
Jorge Zuniga  
Michael McKenna  
Matthew D. Dyer  
Jonathan M. Rothberg

### **ESPCI Paris**

Andrew Griffiths  
Adeline Pichard-Kostuch  
Marco Rubezzi-Crivellari

### **Rutgers University**

Smita S. Patel

### **University of Wollongong**

Antoine M. van Oijen

### **VIB-UGhent**

Nico Callewaert  
Simon Devos