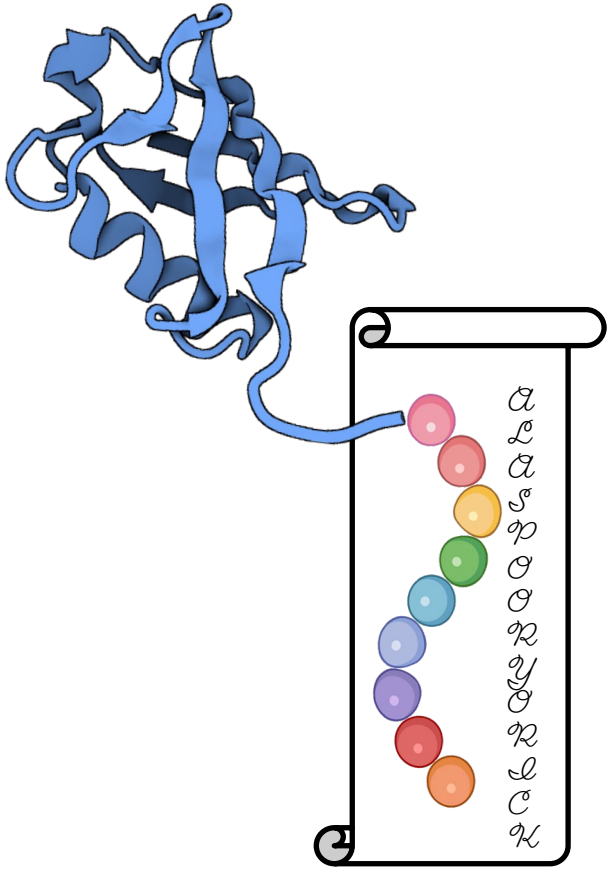
PROSE

PROtein SEquencing

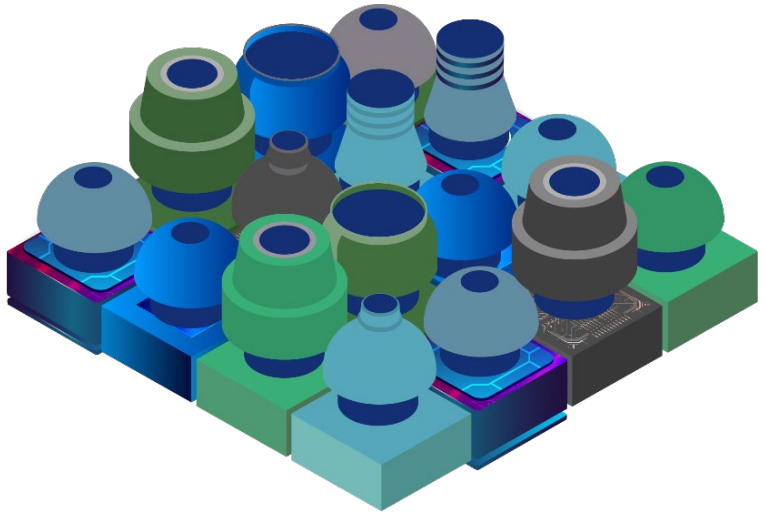
John M. Hoffman, Ph.D.
Program Manager



Industry Day

June 30, 2025

DARPA

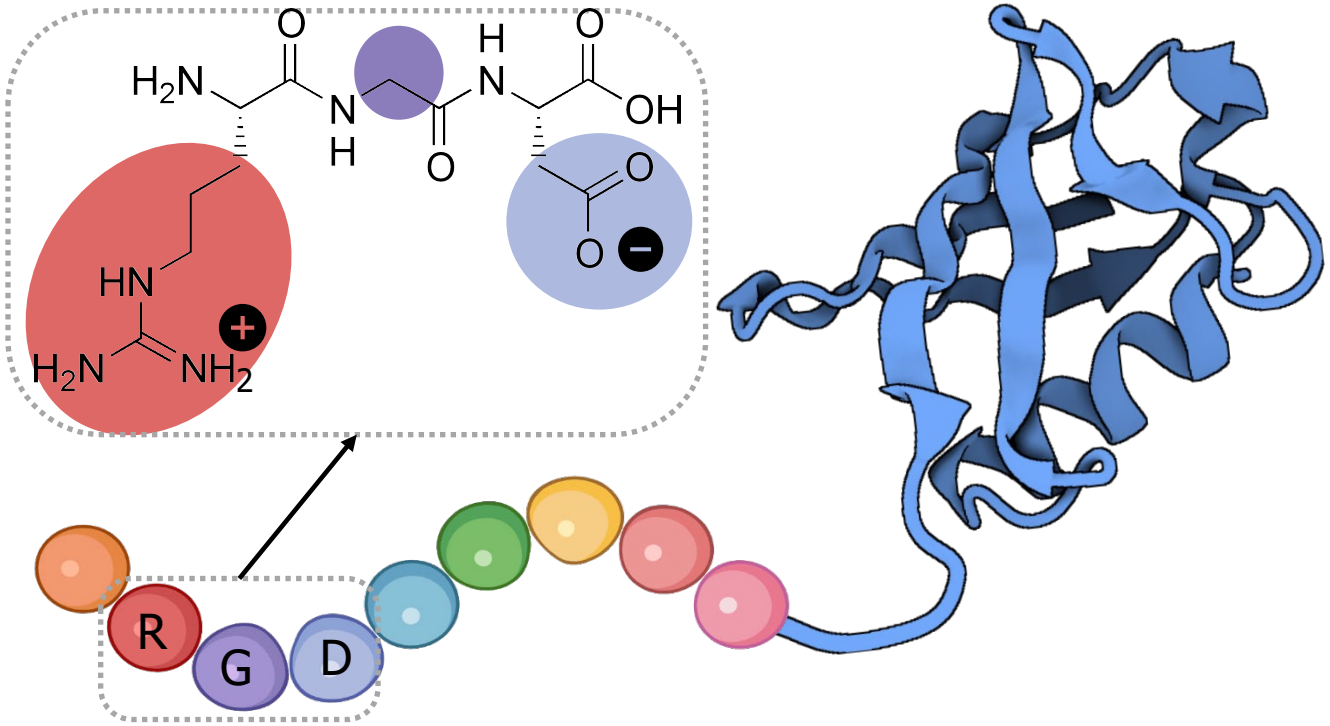


Agenda

PROSE Program Overview

Proteins are complex molecules, and they are difficult to read

Proteins are made of a string of chemically diverse letters, or amino acids and their modifications, and have no uniform charge, shape, or chemistry—which makes them difficult to read



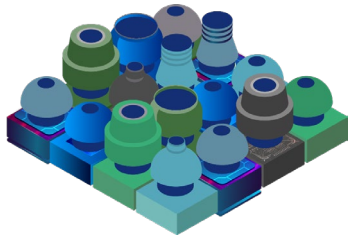
Average protein length:	300
Known letters:	~10^3
"Unknown" letters:	>10^6
Protein read limitation:	<30

As the potential design alphabet is expanding exponentially, our inability to read these sequences without prior knowledge is a vulnerability

Program Goal

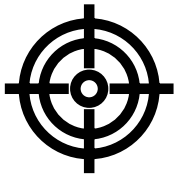
Program goal: Develop a new class of accurate, scalable molecular readers to enable the characterization of unknown protein-based biothreats

One Technical Area



Two Technical Challenges

TC1: Reading complexity at length

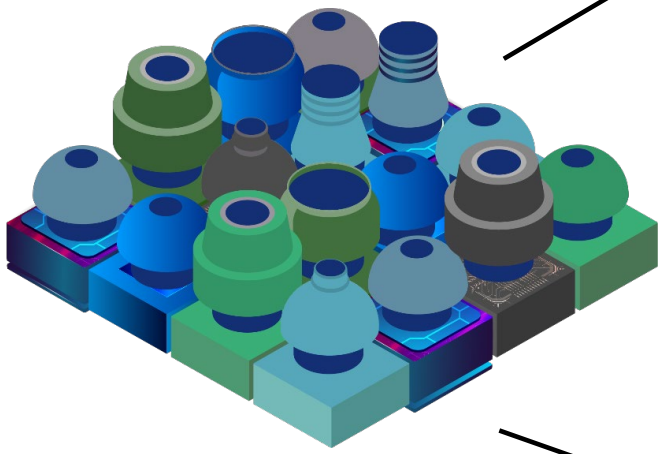


TC2: Reading complexity accurately

Key Insights:

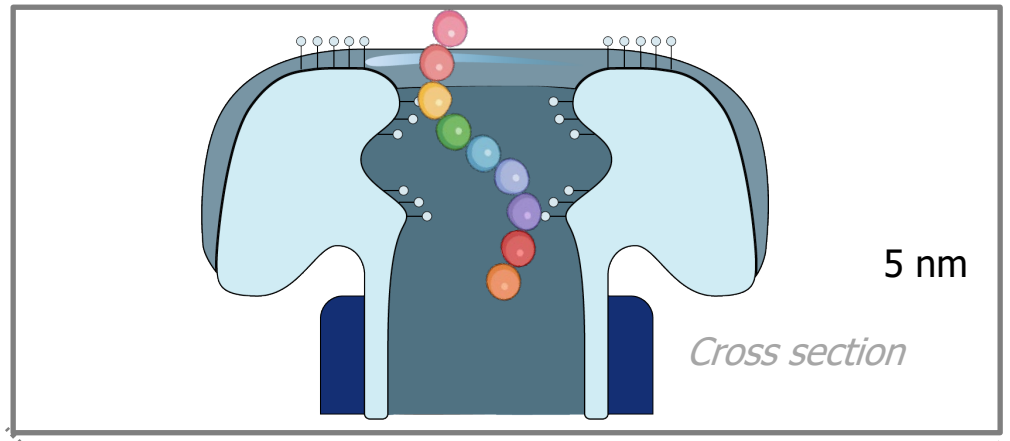
1. Signal transduction approaches leveraging novel chemistry and spectroscopy
2. High resolution read elements integrated within low noise microsystems

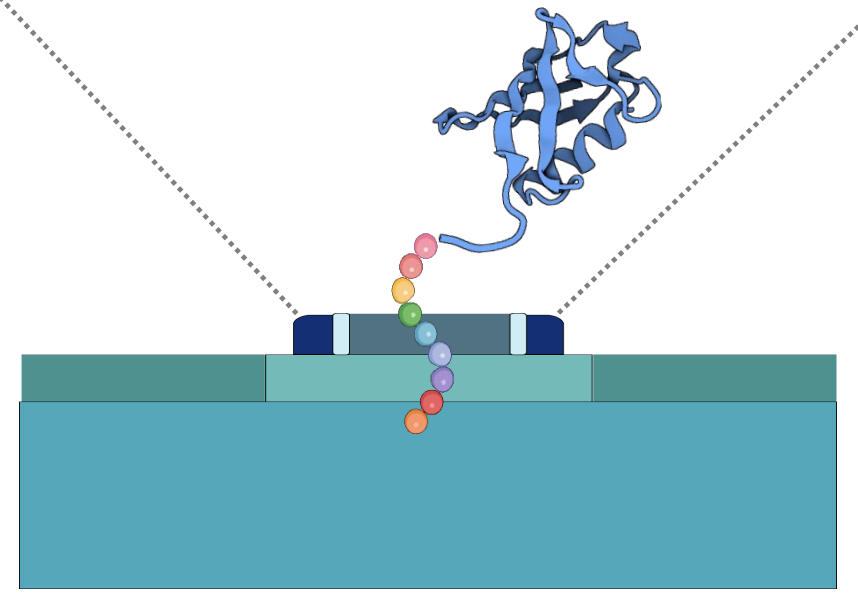
The world's first protein reader



TC1

TC2





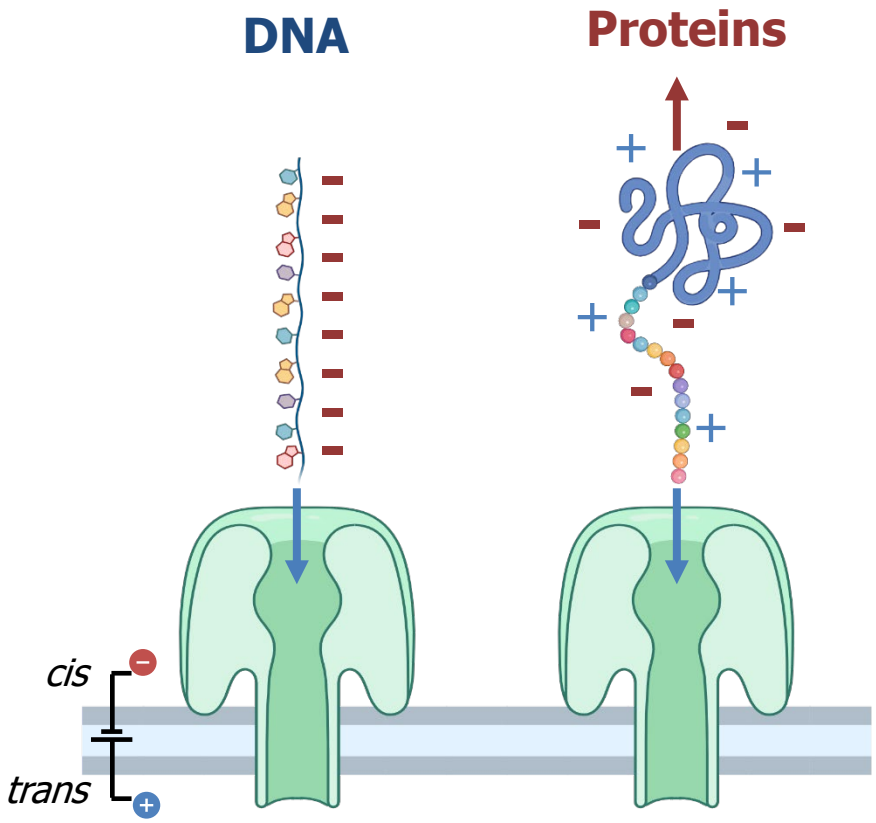
Technical Challenge 1: Reading complexity at length

Technical Challenge 2: Reading complexity accurately

Technical challenge 1: Multi-letter read elements

Sequence length: translocation and movement

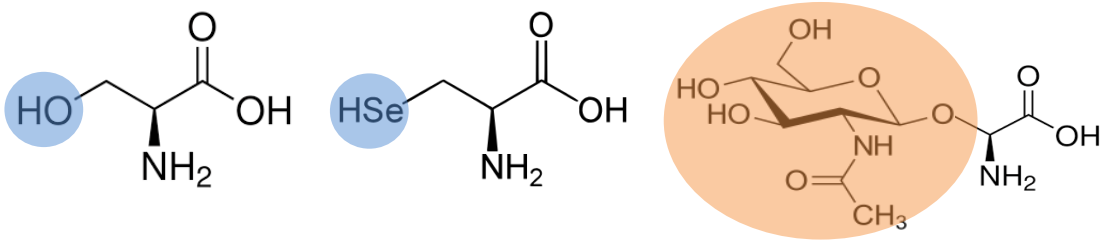
DNA: uniform, negatively charged backbone
Proteins: heterogenous, variably charged



[1]

Alphabet: discrimination between letters

DNA: ~4 letters, chemically similar
Proteins: >10^3 letters, varying polarity, size, charge, composition

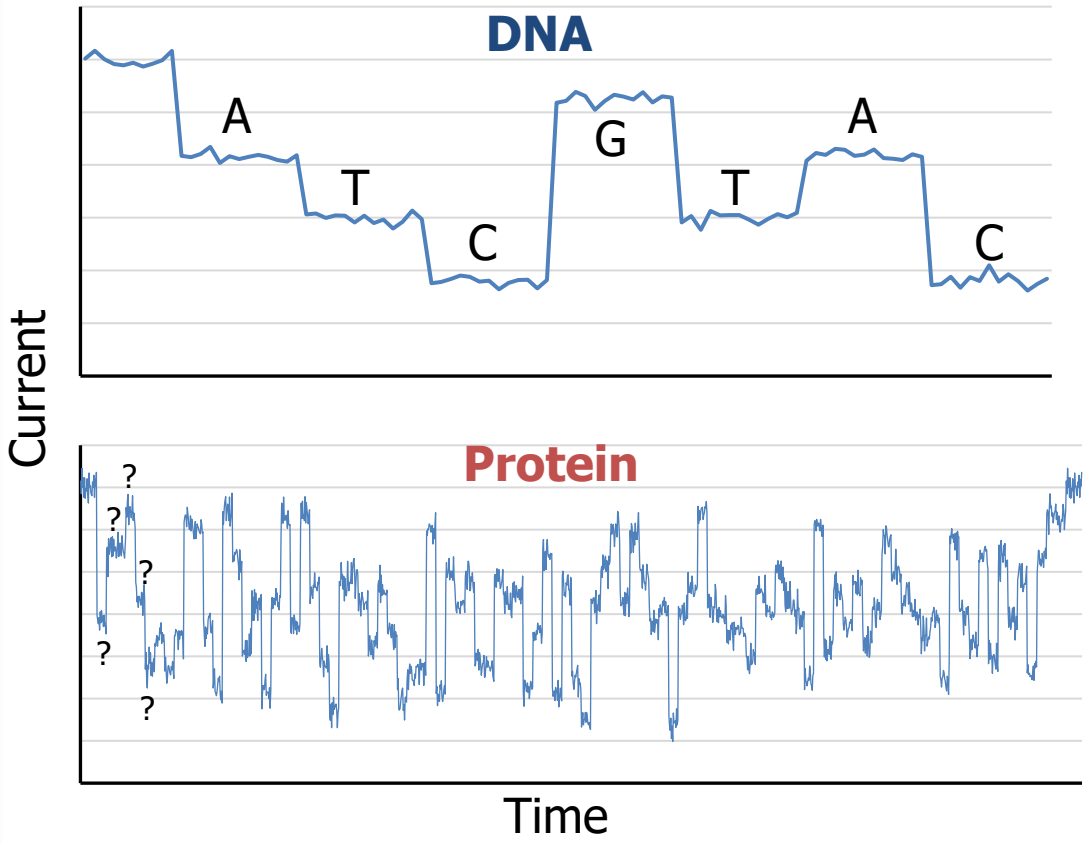


[1] Motone, et al. Nature (2024)

Technical challenge 2: High-accuracy architectures

Resolution: Identifying individual letters

Today: Need to identify 4 letters
PROSE: Need to identify >1000 letters



Noise: More room for more letters

Today: Room for 4 total letters
PROSE: Room for >1000 letters

$$Channel\ Capacity = frequency \times log_2\left(1 + \frac{Signal}{Noise}\right)$$

- More letters requires greater channel capacity
- Microsystems could enable higher frequency and/or increased SNR

Anticipated program metrics*

*Subject to change in program announcement

Metrics	Phase 1 Protein read elements	Phase 2 High accuracy microsystem architectures
Length	≥ 100	≥ 300
Accuracy(1)	99% (simulated)	99% (demonstrated)
Throughput	≥ 10^{10} letters / day (simulated)	≥ 10^{10} letters / day (demonstrated)

(1) Single letter, single read accuracy. Performers must identify each letter in sequence with the associated accuracy for each letter called.

Metric	Phase 1		Phase 2		
	Demo 1	Demo 2	Demo 3	Demo 4	Demo 5
	EOM6	EOM12	EOM21	EOM27	EOM35
Alphabet Size(1)	10(2)	20(2)	50(3)	75(3)	100(3)
Length	≥ 50	≥ 100	≥ 150	≥ 200	≥ 300

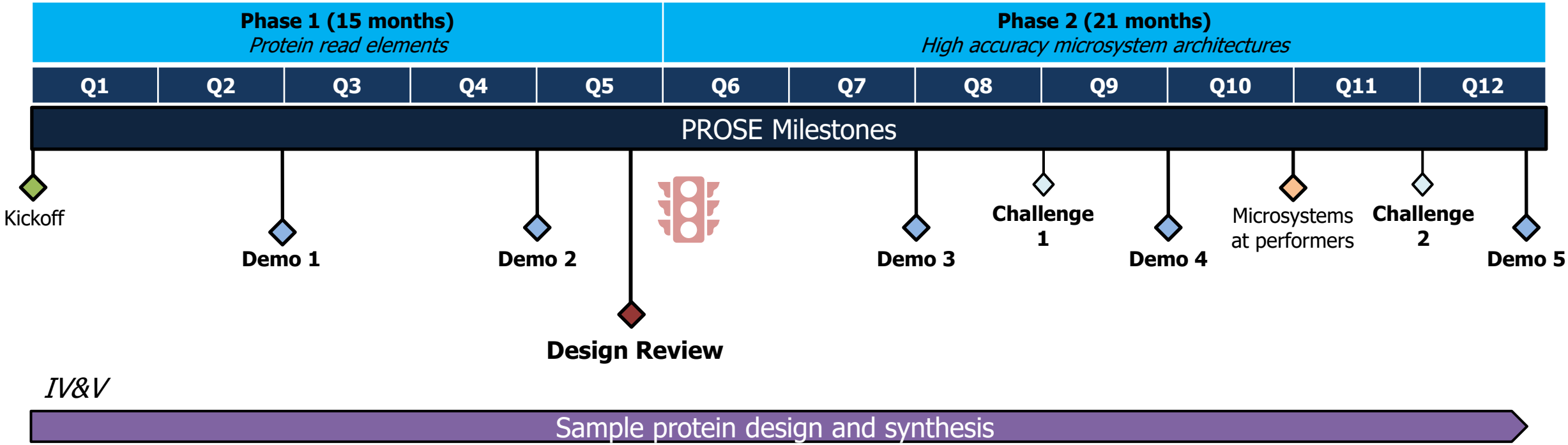
(1) Alphabet consists of natural amino acids, non-natural amino acids, and modified amino acids. Alphabet size is cumulative throughout program (e.g. 50 = 20 previously demonstration + 30 new letters)
(2) The first demonstration will consist of 10 performer chosen natural amino acids, while the second demonstration will cover all 20 natural amino acids.
(3) For each demo, performers will select 10 letters from a list provided by the government team, while the remaining letters will be performer defined.

Notional list of government-defined letters

Phase 1	Phase 2
Arginine	Phosphoserine
Histidine	Phosphothreonine
Lysine	Phosphotyrosine
Aspartic Acid	N-acetyl-lysine
Glutamic Acid	N-acetyl-serine
Serine	N-acetyl-threonine
Threonine	N-acetyl-tyrosine
Asparagine	N-methyl-lysine
Glutamine	N,N-dimethyl-lysine
Cysteine	N,N,N-trimethyl-lysine
Glycine	N-methyl-arginine
Proline	Hydroxyproline
Alanine	Hydroxylysine
Valine	N-formylmethionine
Isoleucine	Citrulline
Leucine	S-palmitoyl-L-cysteine
Methionine	S-farnesyl-L-cysteine
Phenylalanine	S-glutathionyl-L-cysteine
Tyrosine	S-nitrosocysteine
Tryptophan	Nitrotyrosine
	Biotinylated lysine
	O-GlcNAc-Ser/Thr
	Dehydroalanine
	Dehydrobutyrine
	Azidohomoalanine (AHA)
	Homopropargylglycine (HPG)
	p-Azido-phenylalanine (AzF)
	p-Benzoyl-phenylalanine (Bpa)
	Norleucine
	Homoserine

Anticipated program structure*

*Subject to change in program announcement



Demonstrations will be held throughout the program to evaluate improvements read capabilities.

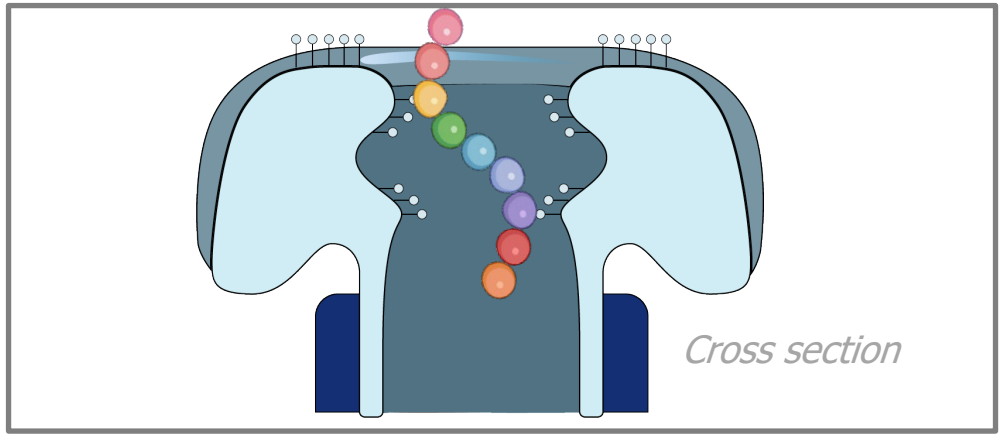
Challenges will be cash prizes awarded to performers that demonstrate read capability beyond program metrics.
For example, reading of additional high impact letters specified by stakeholders.
Challenge specifics will be announced during the course of the program.

Anticipated Phase 1 – Protein Read Elements*

*Subject to change in program announcement

Objective

- Prove the feasibility of reading unknown proteins
- Demonstrate read capability of 20 amino acids for proteins standards
- Simulate the performance of integrated microsystems



Key Metrics and Deliverables

- Two demonstrations of read element performance
- Demonstrate letter calling and accuracy algorithms
- Demonstrate integrated system models that enable the evaluation of Phase 1 accuracy and throughput metrics
- Deliver a microsystem design, including performance simulations and system fabrication/integration approach
- Deliver a commercialization plan

Goals:

- Elements that read 20 amino acids
- Elements that read proteins 100 letters long
- Simulated microsystem performance to achieve 10^{10} letters per day throughput and at least 99% accuracy

Details will be available in the Program Announcement

Anticipated Phase 2 – High accuracy microsystem architectures*

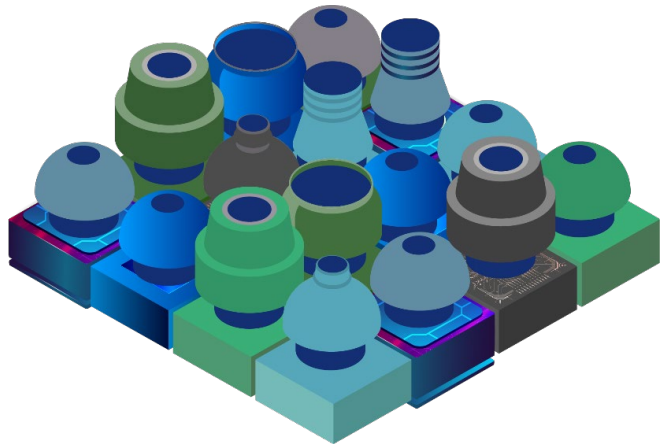
*Subject to change in program announcement

Objective

- Demonstrate high throughput reading of unknown proteins
- Demonstrate read capability of 20 amino acids and at least 80 post translational modifications
- Obtain commercial investment of $50M–100M

Key Metrics and Deliverables

- Two demonstrations of read element performance
- One final demonstration of integrated read element and microsystem performance
- Two challenges for achieving beyond program metrics
- Deliver integrated microsystem
- Deliver letter calling algorithm
- Deliver integrated system model
- Deliver a refined commercialization plan



Goals:

- Systems that read 100 unique letters at 99% accuracy
- Systems that read proteins 300 letters long
- Systems with 10^{10} letters per day throughput

Notes on program structure

- Proposal should address how they will overcome the technical challenges and achieve the program metrics
- It is anticipated the proposers have some preliminary protein read capability. Please include preliminary data and/or publications to support claims.
- Proposals should address the availability and access to any required specialized fabrication capabilities as well as the development of test capabilities
- While not required, DARPA encourages performer teaming as appropriate to achieve program outcomes and commercialize technology
- Proposals should address schedule risks associated with fabrication equipment, test capabilities, and teaming agreements that impact meeting program milestones on schedule
- Proposals should address approaches to meet CUI requirements
- Proposals should address commercialization strategy

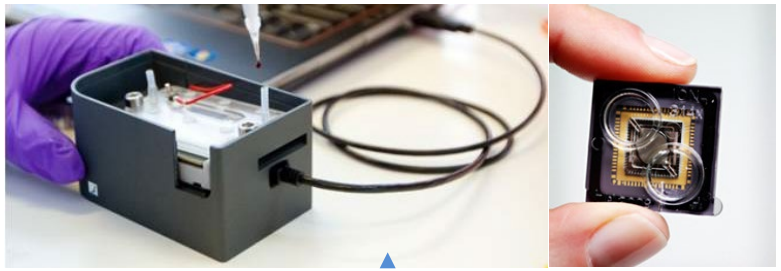
Further details will be available in the Program Announcement

Independent Verification and Validation (IV&V)

- Direct experimental demonstrations of program metrics are required except when specified otherwise
- It is anticipated that demonstrations will take place at performer locations in the presence of government representatives
- It is anticipated that IV&V will synthesize and provide to performers characterized protein standards to assess read capabilities
- It is anticipated that proposers will work with IV&V to develop an IV&V plan that will include, but not be limited to;
  - Synthesizing standards associated with post translational modifications specific to their effort
  - Developing chemistry protocols to adapt standards to the sequencing approach
  - Developing methods to assess simulated performance from integrated system models and microsystem designs

Further details will be available in the Program Announcement

Technology transition and applications



Commercial Investment

Year 3

Year 5

Year 7

PROSE

Industry: ~$1B+

DoD Adv. Dev.

Milestones

Initial comm. plan

Challenge 1

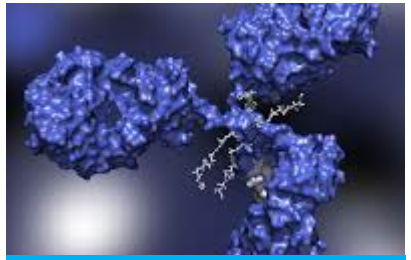
Challenge 2

Refined comm. plan

Unlocked Applications

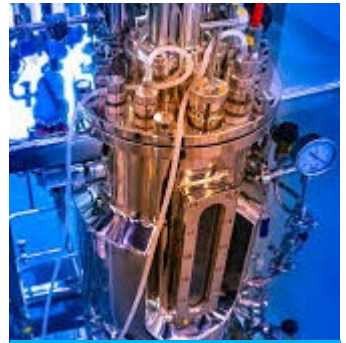
Biosecurity

Diagnostics



Pharmaceuticals

Agriculture



Manufacturing

Remediation

Anticipated PA details

Anticipated proposal guidelines*

*Subject to change in program announcement

- **Abstracts are required** (Attachments A and B)
  - 7-page limit
  - DARPA will provide feedback on whether a full proposal is encouraged or discouraged
  - Allows feedback to proposers on whether a proposed approach or technology fits within the PA
  - Prevents effort spent by proposers on assembling a proposal that may not be successful
- **Proposals** (Attachments C, D, E + all other applicable attachments)
  - Written Content – 25-page limit
    - Expand and refine the Abstract content
    - Provide detailed Statement of Work (or Task Description Document)
    - Include all required attachments; attachments do not count towards the page limit
  - Oral content – Up to 1-hour virtual Q&A
    - Opportunity for the evaluation team to ask clarifying questions
    - **Not an opportunity for proposers to ask the evaluation team questions**
  - Price volume – no page limit

Elements of a Successful Technical Proposal

Goal: Develop a new class of **accurate**, **scalable** molecular readers to enable the characterization of **unknown protein**-based biothreats

- Describe the current state of the art in protein reading including limitations, such as in translocation, discrimination, accuracy, and required bandwidth
- Discuss your current read capabilities and how those enable meeting Phase 1 goals
- Discuss read element and microsystem design, fabrication and integration approaches
- Include analysis of all factors limiting accuracy and scalability
- Discuss the projected accuracy of your proposed read element across a range of chemical complexity
- Discuss the projected scalability of your proposed microsystem to high throughput for long protein lengths
- Describe the computational approaches and algorithms for letter calling, noise modelling, microsystem design and projected performance
- Discuss the fabrication and integration capabilities you would leverage either internally or externally
- Identify and justify the letters you are proposing to read; e.g. high impact, chemically challenging
- Discuss the commercialization strategy and how this aligns with the DoD's need for protein reading technologies

Elements of a Successful Technical Proposal

Goal: Develop a new class of **accurate**, **scalable** molecular readers to enable the characterization of **unknown protein**-based biothreats

- Discuss the underlying (bio)physics that underly your approach
- Include data, from previous experiments and/or modeling to reinforce technical assertions and projections
- Discuss risks associated with your approach and potential mitigations

Questions: **PROSE@darpa.mil**

Anticipated timeline*

*Subject to change in program announcement

- Industry Day: 30 June 2025 (today)
- PA Posting Date: 10 July 2025
- Abstract due date: 24 July 2025 at 1:00 PM ET
- Frequently Asked Questions deadline: XX YYYY 2025 at 4:00 PM ET
- Proposal Due Date: 29 August 2025 at 4:00 PM ET
- Estimated period of performance start: 01 February 2025

Questions: **PROSE@darpa.mil**

Anticipated Evaluation Criteria (part 1)*

*Subject to change in program announcement

1. Overall Scientific and Technical Merit

- Demonstrate that the proposed technical approach is innovative, feasible, achievable, and complete.
- Describe how the proposed approach will achieve each program metric with sufficient detail and supporting experimental measurements, modeling, calculations, and/or simulations.

2. Potential Contribution and Relevance to the DARPA Mission

- Discuss how the proposed effort addresses the DARPA goals of technology transition and facilitates access to PROSE technologies for DoD applications.

3. Budget and Price

- The budget is realistic and accurately reflects the goals and objectives of the solicitation and reflects a sufficient understanding of the level of effort and staffing needed to successfully accomplish the proposed approach.
- It is expected that the effort will leverage all available relevant prior research in order to obtain the maximum benefit from the available funding.
- For proposals that contain resource share, the proposer has provided sufficient rationale as to the appropriateness of the resource share arrangement relative to the objectives of the proposed solution.

Questions: **PROSE@darpa.mil**

Anticipated Eligibility Information*

*Subject to change in program announcement

- DARPA prohibits contractors/performers from concurrently providing Systems Engineering Technical Assistance (SETA), Advisory and Assistance Services (A&AS), or similar support services and being a technical performer, unless the DARPA Deputy Director grants a written waiver

- DARPA extends this prohibition to University-Affiliated Research Centers (UARCs) and Federally Funded Research and Development Centers (FFRDCs) including National Labs, who as a result of their specialized expertise and areas of competencies, are able to accomplish integral tasks that cannot be met by government or contractor resources. Therefore, these entities are **highly discouraged** from proposing against this solicitation as award to a UARC or FFRDC will only be made by exception.

Questions: **PROSE@darpa.mil**

Reminders

- **Submit early**
- It is anticipated that PROSE abstracts will be **mandatory**
- Adhere to the abstract and full proposal guidance and templates provided as PA Attachments
- Corroborate all technical claims with analytical or experimental evidence
- Submit only one proposal per lead organization
- Organizations can be part of multiple proposing teams (as subs)
- When in doubt, refer to the PA for guidance

Questions: **PROSE@darpa.mil**

The PROSE team

- **John M. Hoffman, Program Manager**
  - Taylor Engdahl, Technical SETA
  - Zachary Fishman, Technical SETA
  - Lucas Veillon, Technical SETA
  - Jeff Stone, Programmatic SETA
  - Dann Cameron, Program Security

Questions: **PROSE@darpa.mil**

Join us for the

DARPA MTO Spark Tank

- **Learn about MTO's mission and technology priorities**
- **Engage directly with MTO Program Managers**
- **Explore opportunities to pioneer the next breakthroughs in microsystems**

Register by June 30

July 24-25, 2025 | Aurora, CO