

Technology Snapshot

Technology Licensing Opportunity

SignalGen

A platform that predicts the right signal peptide for a specific protein

Value Proposition

Many medicines, vaccines, and biotechnology products rely on proteins reaching the right cellular compartment or secretion outside a cell. Whether a protein gets there in enough quantities depends on its signal peptide or leader sequence.

Scientists can identify many existing signal peptides, but predicting which one will work best for a specific protein remains difficult. SignalGen uses generative AI to predict the signal peptide most likely to guide a protein to the right location. To make that prediction, it considers the protein, where it needs to go, and the organism it comes from. This gives researchers a new way to improve protein expression for medicines, vaccines, and biotechnology products.

Technology Readiness Level 3

IP Information

S-195368; U.S. Patent pending

Contact Information

licensing@lanl.gov

Overview

Proteins carry out many of the body's essential functions, but they can only do their job if they reach the right place inside or outside a cell. Many proteins rely on signal peptides to get there. Typically, a signal peptide is a short sequence segment at the



Adobe Stock image for illustration purposes only

beginning of a protein that acts like a built-in address label, helping direct the protein to the location where it is needed.

Predicting the right signal peptide for a specific protein remains challenging. Existing software can identify known signal peptides or determine whether a protein already contains one, but it cannot predict which signal peptide is best suited to produce the desired protein expression and cellular location.

Researchers provide the protein they want to study, where they want it to go to a specific compartment in a cell, and the organism. SignalGen predicts the signal peptide best suited for those conditions, giving researchers a stronger starting point before laboratory testing. SignalGen is even critical for AI designed de novo proteins as it would require a signal peptide.

Advantages

- Predicts signal peptides for specific proteins instead of only identifying known ones
- Considers the protein, where it needs to go, and the organism in a single prediction
- Works with both human and non-human proteins
- Natural-language interface reduces the need for programming or scripting
- Demonstrated approximately 90% prediction accuracy using the all-organism model

Continued on reverse

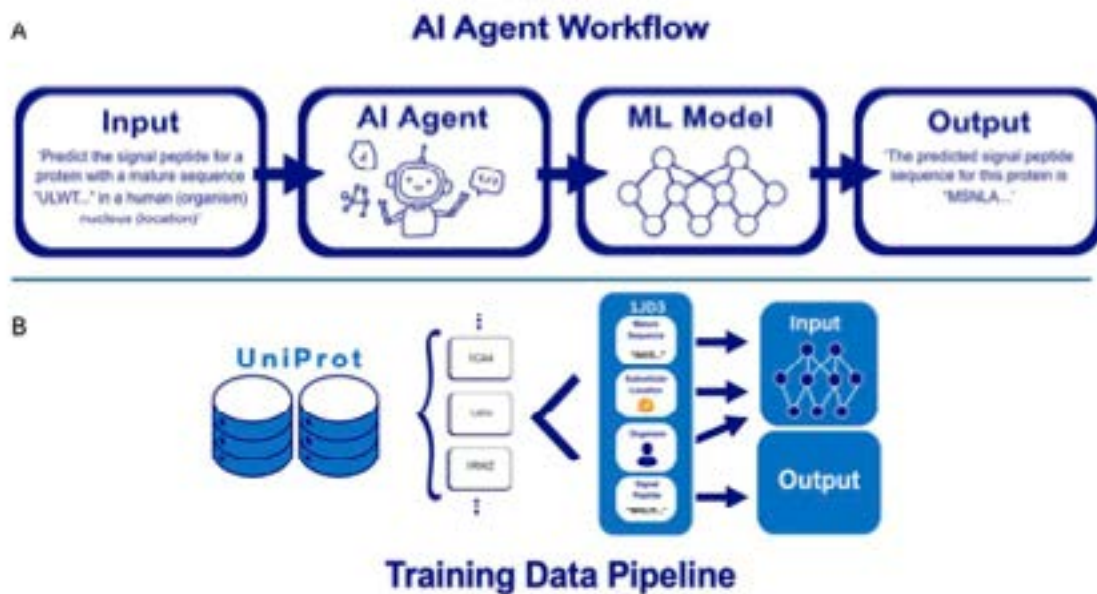


Figure: Schemes of the inference in deployment (A) and preparation of training data (B).

Technology Description

At the core of SignalGen is a protein language model trained on thousands of proteins and their associated signal peptides. By learning the relationships between proteins, signal peptides, where proteins are located inside the cell, and the organisms they come from, the model predicts the signal peptide best suited for a specific protein.

The model combines information about the protein, where it needs to go inside or outside the cell, and the organism it comes from to generate its prediction. It was initially trained using human proteins and later expanded to include proteins from many different organisms, improving its versatility and achieving approximately 90% prediction accuracy.

SignalGen also includes an AI agent that makes the platform easier to use. Researchers interact with the system using natural language rather than programming or scripting. The AI agent gathers the required information, runs the prediction, and returns the recommended signal peptide through a guided workflow.

Market Applications

- Therapeutic protein development
- Vaccine research and development
- Drug discovery
- Biomanufacturing
- Protein engineering
- Synthetic biology
- Industrial biotechnology
- Molecular biology research