CABBI
CENTER FOR ADVANCED BIOENERGY
AND BIOPRODUCTS INNOVATION

Design of Diverse, Functional Mitochondrial Targeting Sequences Across Eukaryotic Organisms using Variational Autoencoder

Background/Objective

Mitochondria play a key role in cellular energy production and metabolism, making them a promising target for metabolic engineering manipulation. However, only a few protein-localization tags that target mitochondria have been characterized. We used a generative artificial intelligence (GenAI) tool called Variational Autoencoder (VAE) to design novel mitochondrial targeting sequences (MTSs).

Approach

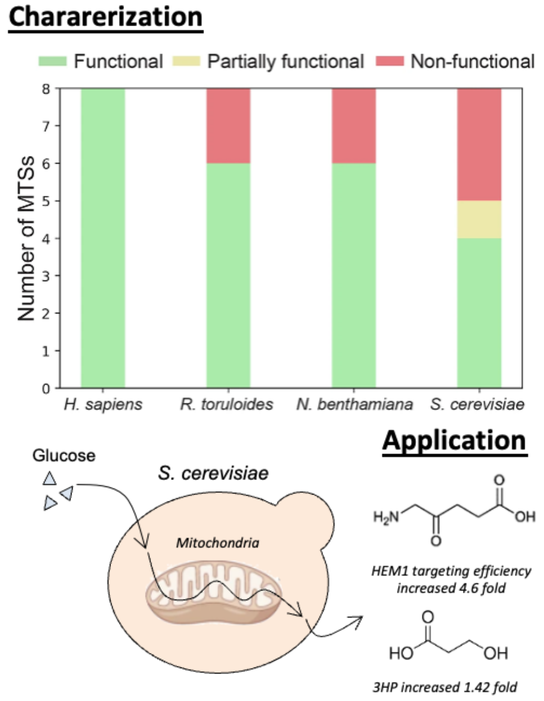
VAE, an unsupervised deep learning framework, was trained on a large dataset of naturally occurring MTSs to generate new-to-nature MTSs, whose functions were experimentally validated in four eukaryotic organisms. Their utility was demonstrated by increasing 3-hydroxypropionic acid (3-HP) acid titers through pathway compartmentalization and improving 5-aminolevulinate synthase (HEM1) delivery in *S. cerevisiae*. Then, a separate VAE was trained to utilize latent space interpolation in designing dual-targeting sequences capable of targeting both mitochondria and chloroplasts, shedding light on their evolutionary origins.

Results

A high fraction of generated MTSs were functional and displayed significant diversity, sharing less than 60% sequence identity with MTSs from the training data. MTSs applied to 3-HP production in mitochondria increased titers by 1.42-fold compared to the cytoplasmic counterpart and improved the targeting efficiency of HEM1 by 4.6-fold.

Significance/Impacts

This work demonstrates the potential of GenAI to design novel, functional MTSs, highlighting their utility in engineering mitochondria for both fundamental research and practical applications.



Characterization of MTSs in four eukaryotic organisms and its application in *S. cerevisiae*.

Boob et al. 2025. “Design of Diverse, Functional Mitochondrial Targeting Sequences Across Eukaryotic Organisms using Variational Autoencoder.” *Nature Communications*. DOI: 10.1038/s41467-025-59499-3.