Title: Redesigning the *Escherichia coli* genome with a 19-Amino Acid Alphabet

Authors: Guillaume Urtecho[1*] (gu2144@cumc.columbia.edu), Liyuan Liu[1], Yiming Huang[1], Charlotte Rochereau[1], Avi Swartz[2], Tomasz Blazejewski[1], Sergey Ovchinnikov[2], Calin Plesa[3], and **Harris Wang**[1]

Institutions: [1]Department of Systems Biology, Columbia University, New York, NY; [2]Knight Campus, University of Oregon, Eugene, Oregon; [3]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA

Website URL: https://sc-programs.llnl.gov/biological-and-environmental-research-at-llnl/secure-biosystems-design

Project Goals:

AIM 1. Computational redesign of all *E. coli* proteins with a 19 amino acid alphabet (Ec19). The overarching goal of this aim is to develop the necessary computational models and designs leading to the generation of functional proteins utilizing a reduced amino acid alphabet (Ec19).

AIM 2. Systematic high-throughput (HT) experimental testing of 19-AA gene designs for all essential and highly expressed *E. coli* genes. The goal of this aim is to develop and implement a HT platform to experimentally test the function of individual gene designs (from Aim 1) directly in cells.

Abstract Text:

The amino acid alphabet of life is universally conserved from bacteria to eukaryotes. As such, all living organisms on Earth require at least 20 amino acids (AAs) to grow and reproduce. This project seeks to answer the following central question: Are all 20 canonical amino acids (AAs) essential for life, or can life be built with fewer than 20 amino acids? Computational analysis of all *E. coli* genes and their orthologs shows that Ile (I), and Val (V) have the lowest frequencies of strong evolutionary conservation. Furthermore, these amino acids exhibit highly similar biochemical properties and are rarely found in active sites of enzymes. Altogether, these observations suggest that Ile or Val may be suitable residues for global substitution. Here, we explore the hypothesis that Ile or Val is dispensable to biological life. We have developed five protein design strategies, ranging in degree of sophistication, to redesign 396 essential or highly-expressed *E. coli* genes where one or both of these amino acids have been globally replaced. These approaches yield 1606 redesigned variants, which we have synthesized using DropSynth and uniquely barcoded for downstream characterization in pooled formats. Lastly, we perform the direct replacement of native *E. coli* genes with their cognate variants using a lambda recombinase-based approach and characterize the viability of redesigned genes in multiplexed growth assays. We anticipate these experiments will generate a rich set of data that can be used to improve and refine protein models developed in Aim 1 and to provide a platform for iterative design improvements for all genes in *E. coli*.

Funding Statement:
This work is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Lawrence Livermore National Laboratory Secure Biosystems Design SFA "From Sequence to Cell to Population: Secure and Robust Biosystems Design for Environmental Microorganisms", grant no. 20-1777; the National Science Foundation, Grant no. MCB-2032259; and the Howard Hughes Medical Institute Hanna H. Gray Fellows Program, grant no. GT15182.