Cell Development Kit: E. coli – Small Molecules: Metagenomic Discovery

MODULE OVERVIEW

SUMMARY

- Create a metagenomic enzyme library for improved activity in a specific target reaction, followed by screening in *E. coli* to quantify enzyme performance and identify top candidates.
- Validate the genotype of the top-performing *E. coli* strains through NGS and deliver the top 10 enzyme sequences meeting or closest to the activity goals.

RESOURCES

- CDK Library (Duration/Cost/Teams Involved)
- Instructions for Copying Standard Module Smartsheet Template Rows into Project Plans
- Technical Work Plan
- @Thomas Kakule [Tech Lead]
- @Hui Zhou [Senior Program Lead]

ABSTRACT

This module streamlines enzyme discovery for specified reactions and substrates by sourcing homologous sequences from metagenomic databases. Discovery for enzyme activity on a new substrate or for identifying homologs with higher activity can be addressed by running this module. The workflow involves library design, *E. coli* transformation, and high throughput screening. In vitro assays are typically used, but in vivo assays are also accepted if in vitro screening is not an option.

MODULE INTRODUCTION

What It Does:

- Leverages our proprietary 2-billion metagenomic sequence database and the latest sourcing techniques developed by the protein engineering team to design a library for the specific target reaction.
- Screens the library with an *E. coli* plasmid expression system to quantify the activity of enzymes for the specific target reaction. The activity readout can be from an in vitro assay or an in vivo-based plate screen.

How It Works:

- Starts with designing and synthesizing a metagenomic-sourced library *without a tag*.
- Utilizes high-throughput methods for expression and screening of enzymes. A process control validates the reaction conditions for the screen.
- Identifies sequences with the enzyme activity of interest through systematic evaluation, and ranks them relative to a positive control when such a control is available.
- Validates the selected top strains through next-generation sequencing.

When to Use It:

- When there is a need to discover enzymes with activity on new substrates using a metagenomic-sourced library.
- For identifying more active enzyme homologs for a reaction of a biosynthetic pathway in *E. coli*.

- When in vitro assaying for the reaction of interest is not an option, e.g. whole-cell biotransformation solutions for living therapeutics or membrane proteins.

BASES & ADD-ONS

Base	
Base description	**When to use**
Tests a 500-member metagenomic sourced library, in an *E.coli* strain using a high-throughput deep-well plate format	Use this module to identify homologs for the reaction of interest, starting from a seed sequence

ADD-ONS		
Add-On Name	**Optional or Required**	**Description**
Secondary Hit Confirmation or Orthogonal Assay	Optional	This add-on could be optionally used after the Base when the assay does not support the scale to allow for sufficient replication or when an orthogonal assay/down-selection criteria has to be employed. A popular orthogonal assay is Michaelis-Menten in vitro characterization of hits.
ambr250 Evaluation (Cryo-vialing included)	Optional	Validates the top 10 hits in one process relevant to the expression system at ambr250 fermentation scale. This add-on is not a typical request for this enzyme discovery module.
Material generation via 5/10/30L	Optional	Generation of material from one 5/10 or 30L run, using the identified top hit. Yields are titer dependent.

INPUTS & OUTPUTS

INPUTS
Ginkgo Records (if available): • Index of links to previous Ginkgo data and reports related to the project
Customer Requirements Document and Materials: • One or more seed Amino acid sequences for metagenomic sourcing. Multiple seeds must share >60% identity. • Desired enzyme substrate(s) and product(s). Include CAS numbers of all non-enzyme molecules involved in the reaction. • Cofactors needed for enzyme activity, if applicable. • Metric for scoring enzyme activity. • Customer enzyme or functional assay details, if available. • Physical and chemical properties of substrate, product, and cofactors. • Relevant literature.
Enzymatic Activity Assay Requirements:

OUTPUTS
Report on the library of enzymes • Metagenomic protein-sourcing library design details. • Expression backbone, promoter, RBS and codon-optimization encoding. • Synthesis and assembly success details of the DNA library. • Details of high-throughput transformation of the DNA library into *E. coli*.
Report on HTS library screen results: • Results of high-throughput screening for enzyme activity. • Identification of top strains for further validation. • Next-generation sequencing results for top strains.
Report Generation: • Executive summary and candidate sequences and strains list. • Data supporting sequence identification and any protein characterization. • Suggestions for project next steps.

- Assay principles for direct or indirect measurement of enzyme activity.
- SOP for measuring enzyme activity.
- Mass spectrometer-based method for titer measurement, if needed.
- Kinetic assay SOP, if Michaelis-Menten characterization add-on is requested.

MODULE INFORMATION

TECHNICAL METHODS

Key Inputs : One seed sequence and an analytical SOP supporting screening of 3000-6000 samples at 96-deep-well or 384-well plate scale.

Base Task 1 : Design of a metagenomic protein-sourcing library and DNA synthesis

`Library Size` : 500 - 2000

`Expression System and vector` : T7 expression system without a tag (Default)

Base Task 2 : High-throughput transformation in an *E. coli* host strain and lysate-based plate screening

`Host Strain` : BL21 (AI)

`Deep-well plate cultivation protocol` : Standard 2 day process with arabinose and IPTG induction

`Assay method` : Product-specific functional assay (spectrophotometric reading or/and MS-based analytics)

Add-on Task 3 : Hit-validation via a secondary confirmatory lysate-based screen (add-on)

`Percent library re-screened` : 10% in technical triplicate.

`Assay method` : Product-specific functional assay (spectrophotometric reading or/and MS-based analytics)

GENETIC PARTS (Preliminary guidance based on [WIP] [RFC] CF-RFC-00X: Prefab designs in Loom)

- The library design should follow the Prefab design v. 0.1.0 (as shown in the Design 15604 template) unless a case is made for alterations:
  - Backbone: pG9m-2_Kan_BsmbI cleaned_LacI (DU 216814)
  - Promoter: pT7 ; Operator | lacO; derived from m915111 (DU 636063)
  - RBS: T7_rbs; derived from m915111 (DU 636066)

  A copy of sfGFP on prefab design 0.1.0 (m6382597) should always be ordered as a negative process control.

  If the library design differs from 15604, a sfGFP should also be included as an in-library, benchmark control
- For in vivo characterization within a pathway context, consider using:
  - low-copy vector : pSC101-Kan recommended (backbones m6059762 (Twist), and m2523482 (Biofab))
  - promoter and RBS : identified from assay dev with seed enzyme (BBa_J231x series promoters are recommended)
- No-tags Allowed

Technical Tools

- Primer to Ginkgo Enzyme Sourcing Capabilities (UMDB)
- Consult with PE team, if no seed is available
- Details for the PE team enzyme sourcing offering

Case Studies / References

- Geno programs
  - P156
  - P229

TECHNICAL TOOLS

- Codon-Optimization
- Signal peptide prediction

PROTOCOLS

- In vitro assays utilize standard Bugbuster lysing protocol
- In vivo assays: DWP culturing SOP (developed with project ladder strains)

MODULE DETAILS

MODULE COMPONENT	*DBTL SEGMENT*	Offerings and foundry teams (foundry catalog)	Interactions with other modules (dependencies) and add-ons	Controls	Quality plan and gates	Limitations
BASE	*DESIGN*	`Protein Engineering` - Protein Sourcing Library Design and Transaction Placement		1. Seed sequence 2. Negative control: GFP strain		
	BUILD	Twist DNA Synthesis > 3.2 kb			>90% fulfillment for test library; 100% fulfillment for controls	
		`Ops` Transformation - Production Transformation - High Throughput Transformation			1. Minimum of 2 clones stocked for >90% of the library attempted 2. 10% of library submitted for sequencing QC	Does not accept low-efficiency transformation E. coli hosts or slow-growing strains
		`NGS` - Cellfie			QC check - Right library transformed into correct strain; correct controls; no sample mix-up	
	TEST	`HTS` - Microbial - HTS - Screen Execution		1. *E. coli* strain expressing seed control sequence(s) OR 2. Lysate of *E. coli* strain expressing gfp plasmid	Refer to Gen. SOP	

				spiked with reaction product standard		
		`OAC` - Placeholder	Analytical Method for Product	Supernatant spiked with reaction product standard; in-plate calibration curve	Refer to analytical SOP	
		`NGS` - Cellfie			QC check - Confirm sequences of hits match LIMS info	
	LEARN	`Protein Engineering` - PE - Data Analysis				
Add-on: Secondary Confirmatory screen	*TEST*	`HTS` - Microbial - HTS - Screen Execution	Up to 10% of top hits from Base re-evaluated in a secondary screen.	1. *E. coli* strain expressing seed control sequence(s) OR 2. Lysate of *E. coli* strain expressing gfp plasmid spiked with reaction product standard		
		`OAC` - Placeholder (if needed)	Analytical method for product quant	Supernatant spiked with reaction product standard; in-plate calibration curve	Refer to analytical SOP	
		`NGS` - Cellfie			QC check - Confirm sequences of hits match LIMS info	
	LEARN	`[Codebase]` Review of data				
Add-on: Ambr250 Validation	*TEST*	`NGS` - Cellfie	Sequencing cryovials of hits from Base		QC check - Confirm sequences of hits match LIMS info	
		`Fermentation` - Ambr250 Run	Up to 10 hits from Base evaluated in ambr250 fermentation			

		`OAC` - Placeholder - LCUV	Fermentation Monitoring service		Refer to s19 service SOP	
		`OAC` - Placeholder	Analytical method for Product quant		Refer to analytical SOP	
	LEARN	`[Codebase]` Review of data				