Dissertation Defense

Design of Modular Cell Systems for Biocatalysis with Multi-Objective Optimization

Sergio Garcia

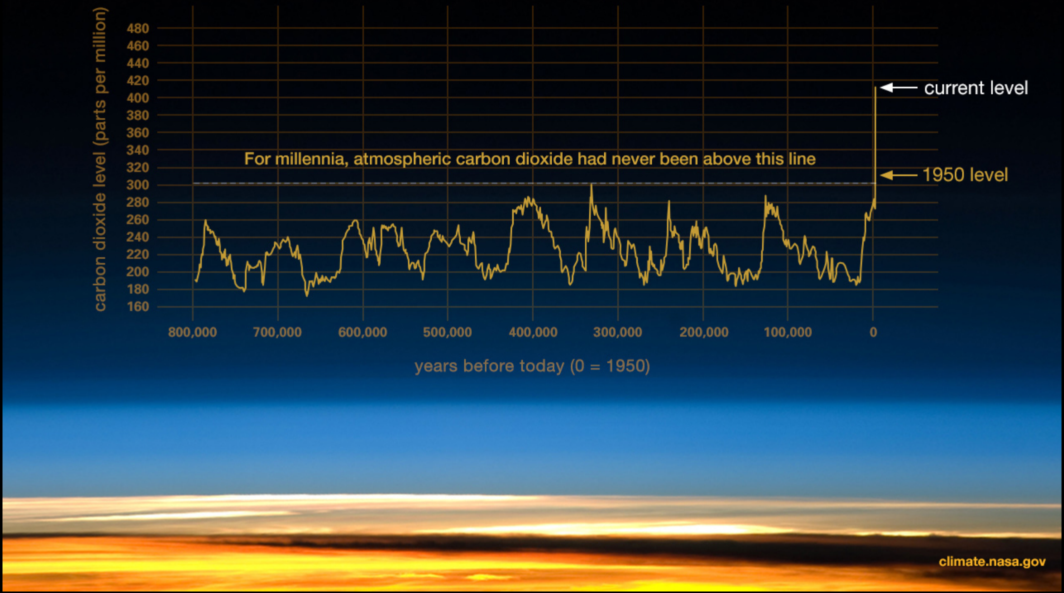
Committee members:

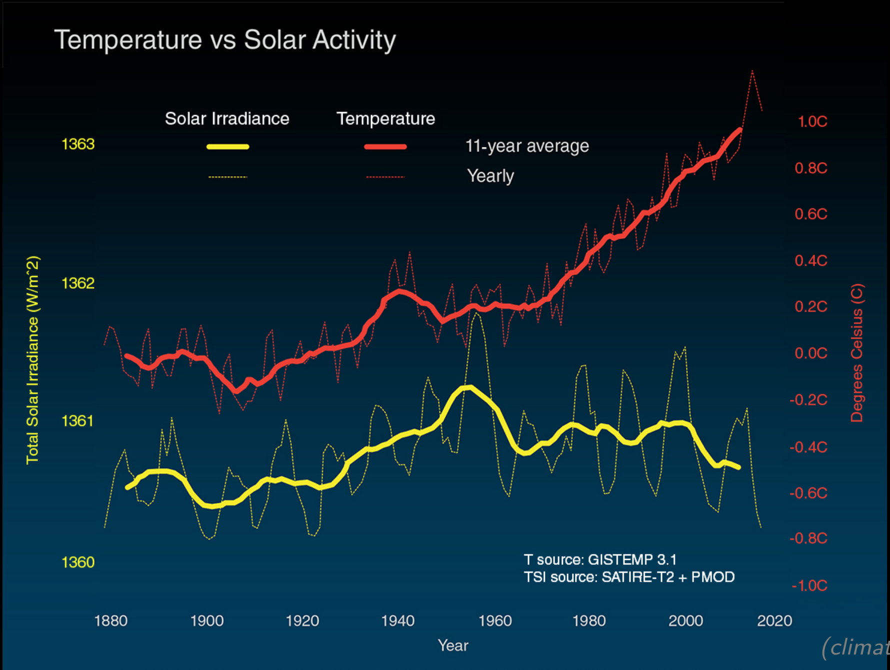
Dr. Steven M. Abel
Dr. Brian H. Davison
Dr. Tian Hong
Dr. Cong T. Trinh (*Advisor*)

Department of Chemical and Biomolecular Engineering
University of Tennessee Knoxville

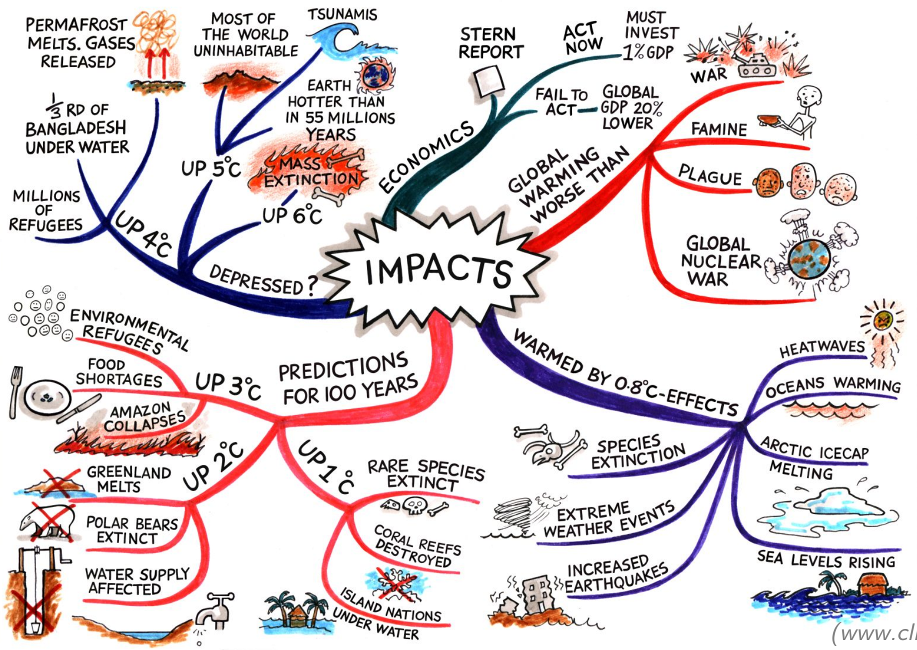
March 25th, 2020

Outline


carbon dioxide level (parts per million)
480
460
440
420
400
380
360
340
320
300
280
260
240
220
200
180
160
current level
For millennia, atmospheric carbon dioxide had never been above this line
1950 level
800,000
700,000
600,000
500,000
400,000
300,000
200,000
100,000
0
years before today (0 = 1950)
climate.nasa.gov

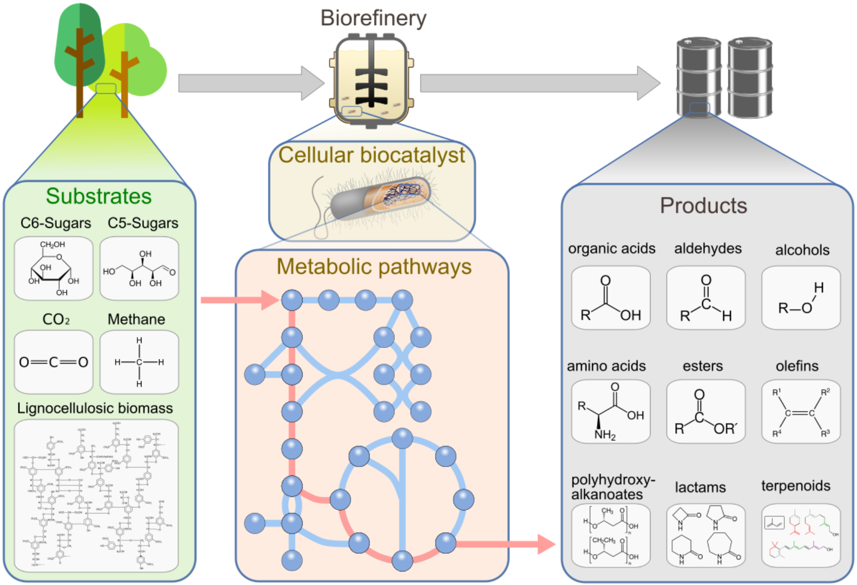


(climate.nasa.gov)



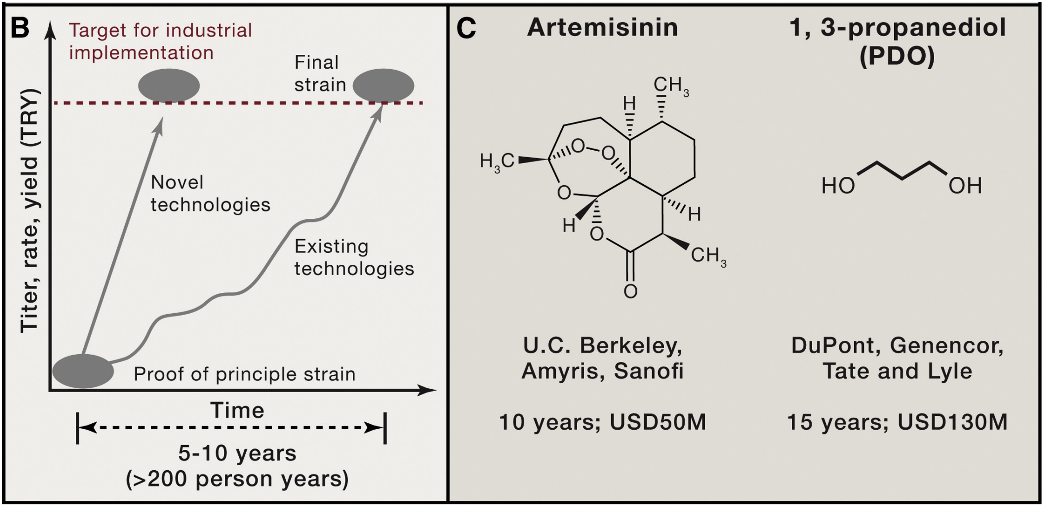
(www.climate.gov)

Cell biocatalysis technology: The goal



- Reduce net CO_2 emissions
- Facilitate decentralized manufacturing

Cell biocatalysis technology: Current state



(Nielsen and Keasling 2016)

Outline

Modular design concepts

Definition:

> "A *module* is an essential and self-contained functional unit relative to the product of which it is part. The module has, relative to a system definition, standardized interfaces and interactions that allow composition of products by combination."

Modular design concepts

Definition:

> "A *module* is an essential and self-contained functional unit relative to the product of which it is part. The module has, relative to a system definition, standardized interfaces and interactions that allow composition of products by combination."

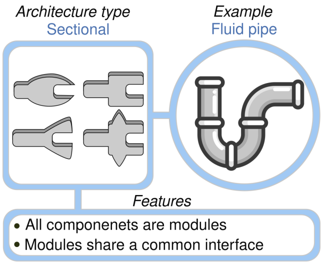
Types of modular architecture:

Modular design concepts

Definition:

> "A *module* is an essential and self-contained functional unit relative to the product of which it is part. The module has, relative to a system definition, standardized interfaces and interactions that allow composition of products by combination."

Types of modular architecture:

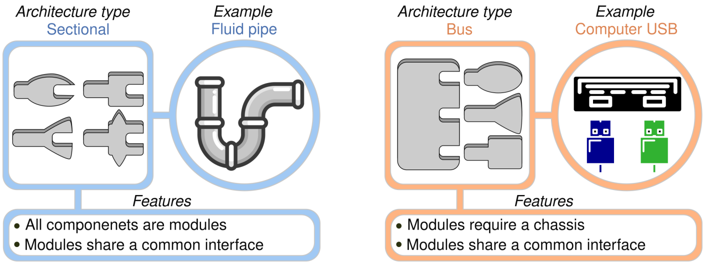


Modular design concepts

Definition:

"A *module* is an essential and self-contained functional unit relative to the product of which it is part. The module has, relative to a system definition, standardized interfaces and interactions that allow composition of products by combination."

Types of modular architecture:

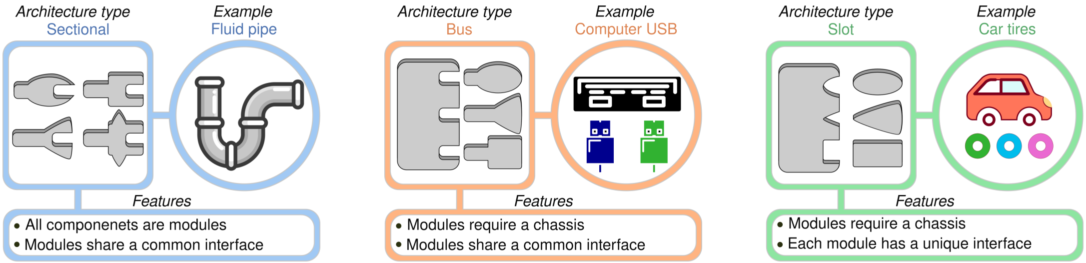


Modular design concepts

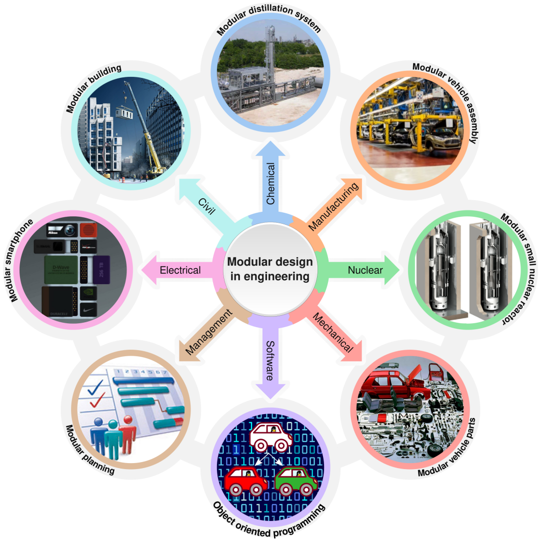
Definition:

"A *module* is an essential and self-contained functional unit relative to the product of which it is part. The module has, relative to a system definition, standardized interfaces and interactions that allow composition of products by combination."

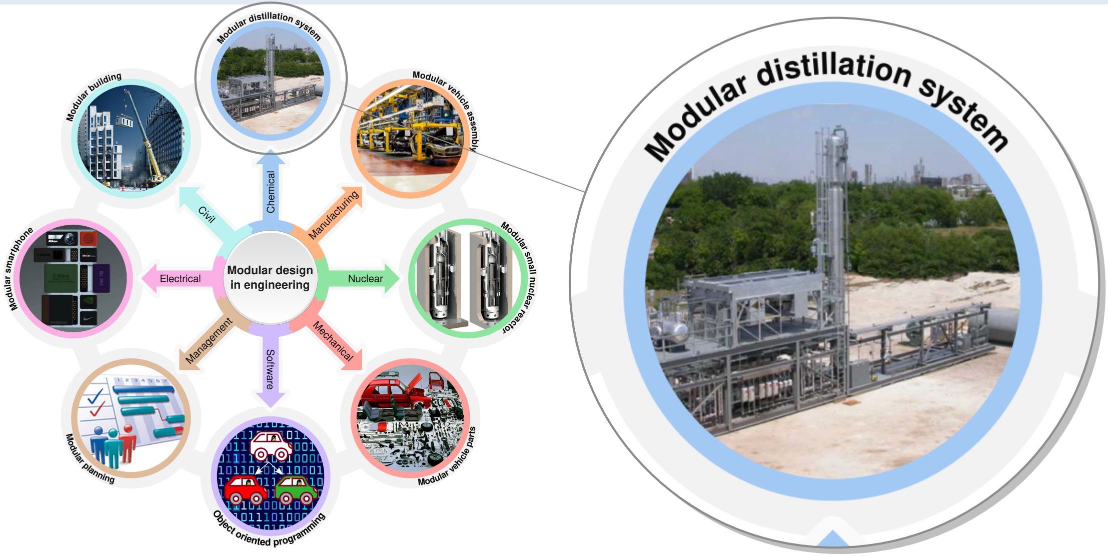
Types of modular architecture:



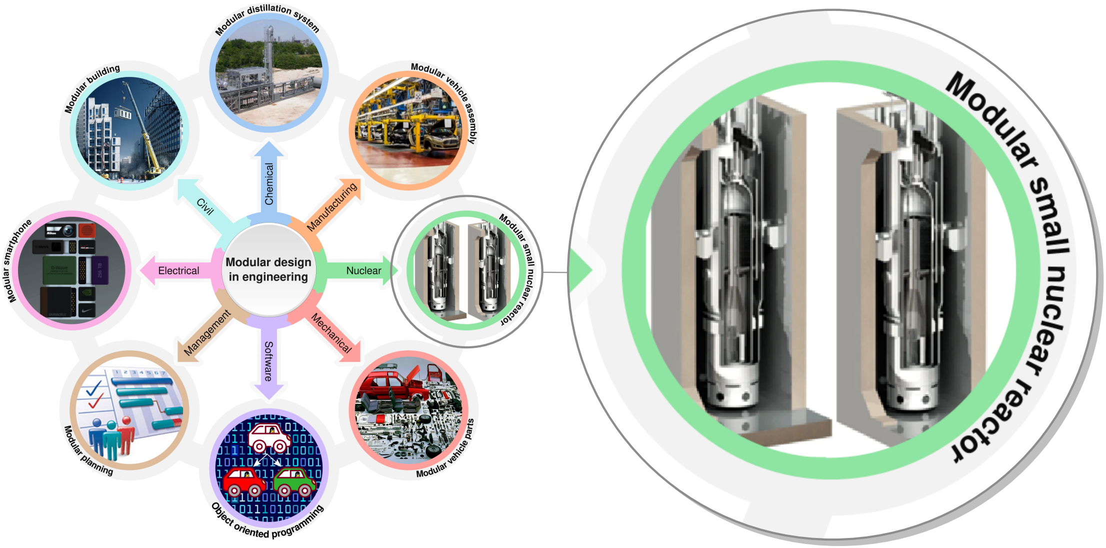
Driving forces and potential tradeoffs of modular design



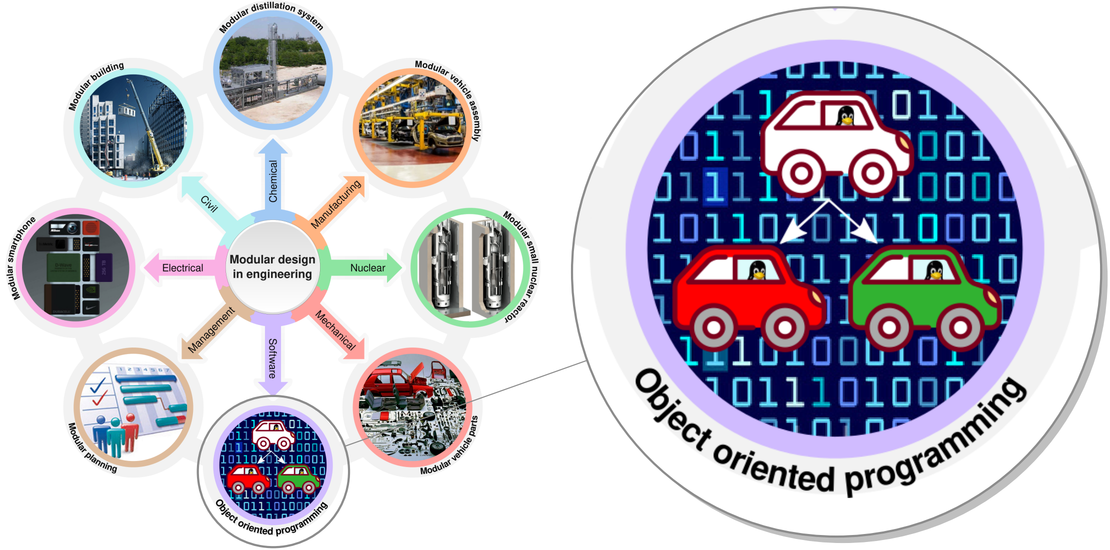
Driving forces and potential tradeoffs of modular design



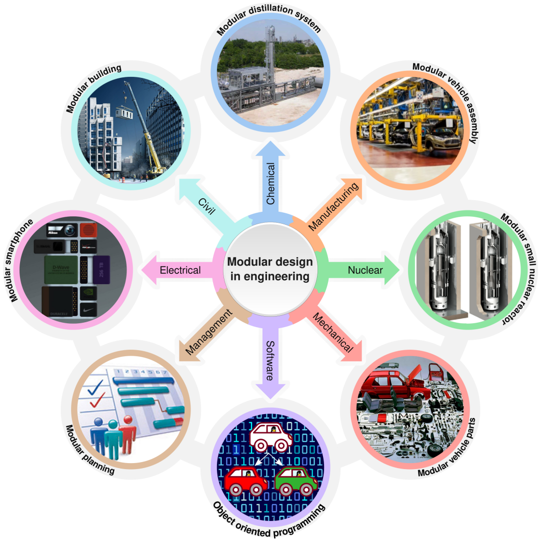
Driving forces and potential tradeoffs of modular design



Driving forces and potential tradeoffs of modular design

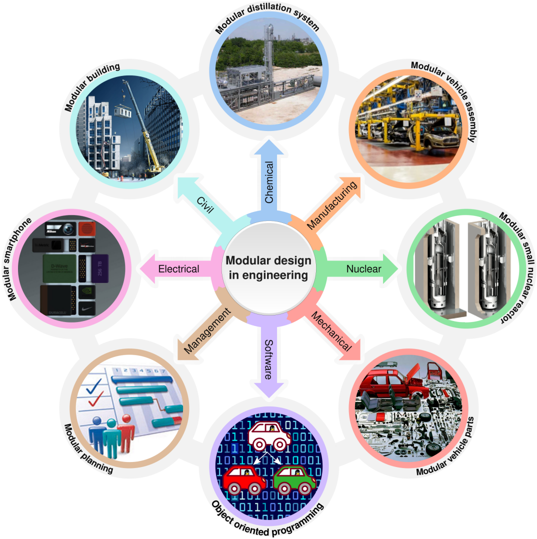


Driving forces and potential tradeoffs of modular design



Driving forces for modularization:

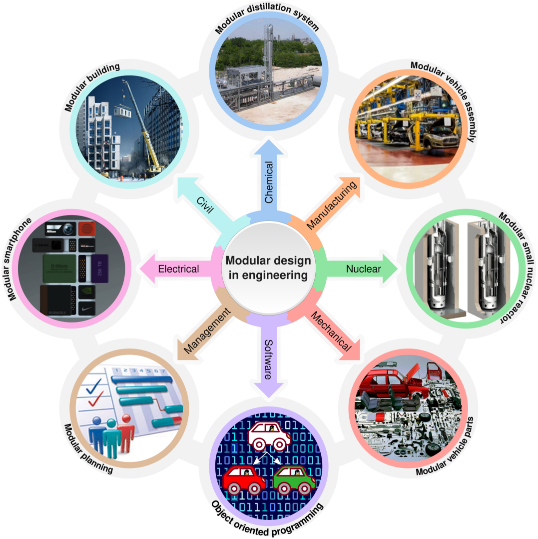
Driving forces and potential tradeoffs of modular design



Driving forces for modularization:

- **Innovation**: Novel solutions to existing problems

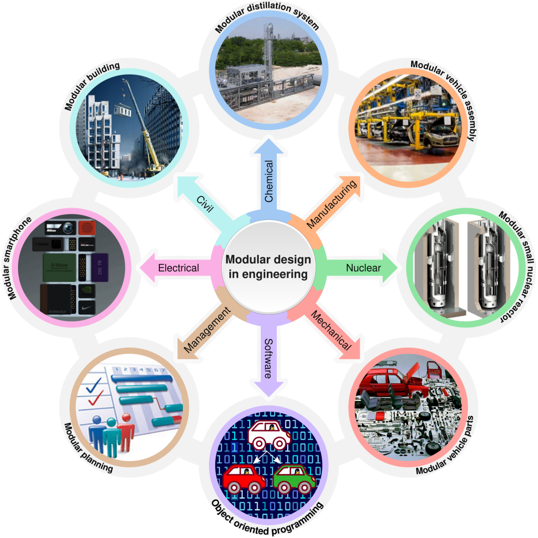
Driving forces and potential tradeoffs of modular design



Driving forces for modularization:

- Innovation: Novel solutions to existing problems
- Efficiency: Faster and cheaper product construction and maintenance

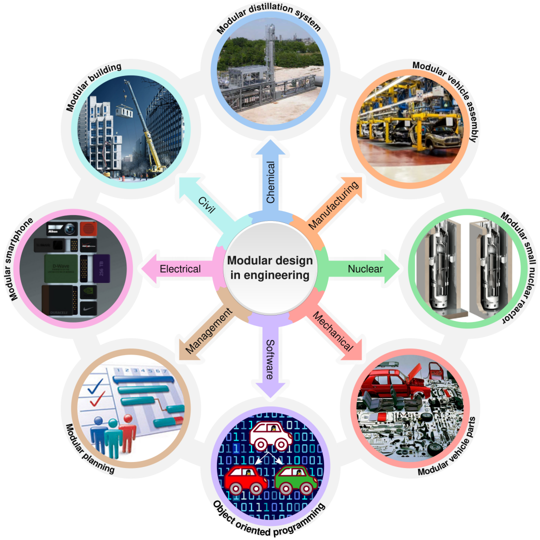
Driving forces and potential tradeoffs of modular design



Driving forces for modularization:

- **Innovation**: Novel solutions to existing problems
- **Efficiency**: Faster and cheaper product construction and maintenance
- **Customizability**: Better tailor a solution to specifics of the problem

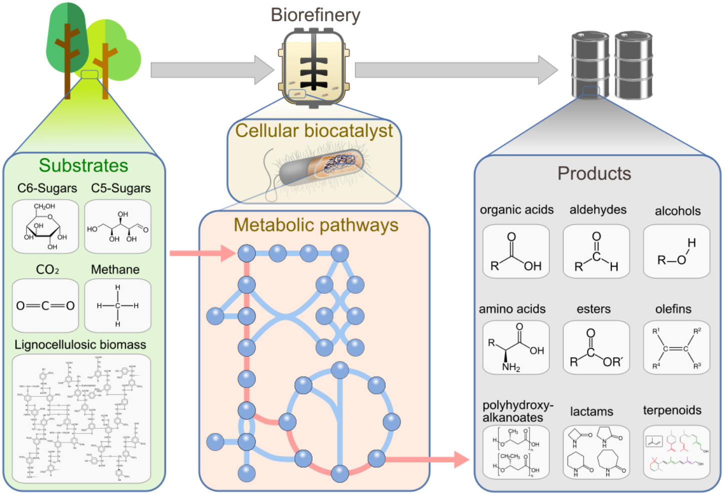
Driving forces and potential tradeoffs of modular design



Driving forces for modularization:

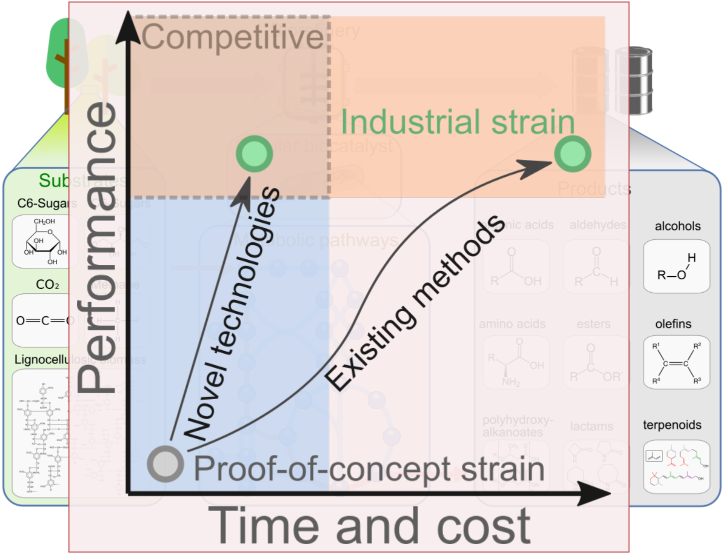
- Innovation: Novel solutions to existing problems
- Efficiency: Faster and cheaper product construction and maintenance
- Customizability: Better tailor a solution to specifics of the problem
- Predictability: Robust system behavior across diverse scenarios

Applications and challenges in microbial catalysis



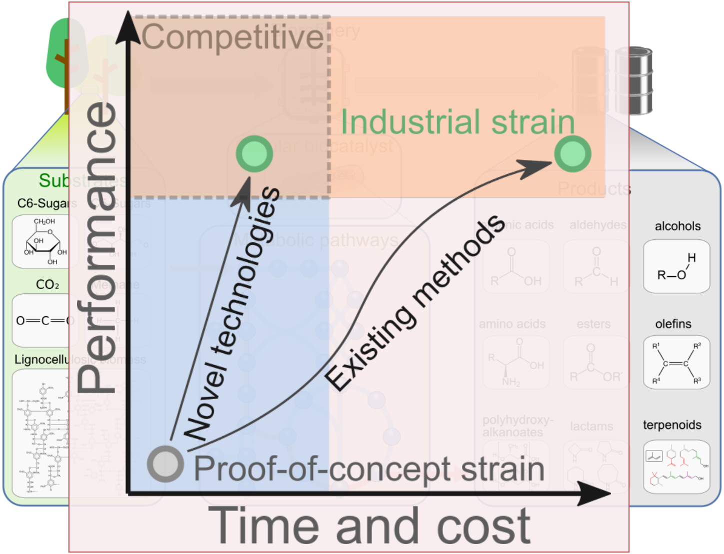
- Status: Bioengineering technologies enable microbial biocatalysis

Applications and challenges in microbial catalysis



- Status: Bioengineering technologies enable microbial biocatalysis
- Challenge: Design-build-test cycles are too slow to make biocatalysis widely applicable

Applications and challenges in microbial catalysis



- Status: Bioengineering technologies enable microbial biocatalysis
- Challenge: Design-build-test cycles are too slow to make biocatalysis widely applicable
- Solution: Apply proven modular design principles to biocatalyst engineering

Modular design in synthetic biology

The second wave of synthetic biology: from modules to systems

Priscilla E. M. Purnick & Ron Weiss

Nature Reviews Molecular Cell Biology **10**, 410–422(2009) | Cite this article

Modular design in synthetic biology

The second wave of synthetic biology: from modules to systems

Priscilla E. M. Purnick & Ron Weiss

Nature Reviews Molecular Cell Biology **10**, 410–422(2009) | Cite this article

The first wave:

- Basic elements (e.g., rbs, promoters, repressors) combined to form small modules (e.g., Switches, oscilators, logic formulas.)
- Modules can be used to regulate gene expression, protein function, metabolism, and cell–cell communication.

Modular design in synthetic biology

The second wave of synthetic biology: from modules to systems

Priscilla E. M. Purnick & Ron Weiss

Nature Reviews Molecular Cell Biology **10**, 410–422(2009) | Cite this article

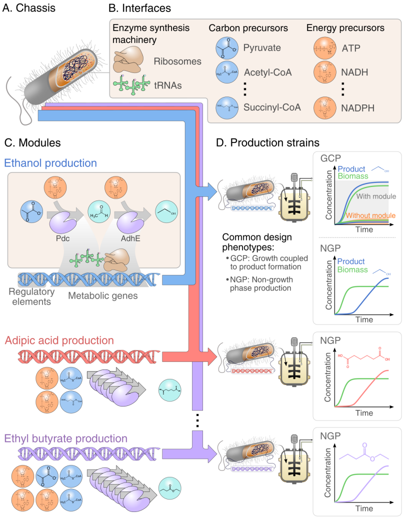
The first wave:

- Basic elements (e.g., rbs, promoters, repressors) combined to form small modules (e.g., Switches, oscilators, logic formulas.)
- Modules can be used to regulate gene expression, protein function, metabolism, and cell–cell communication.

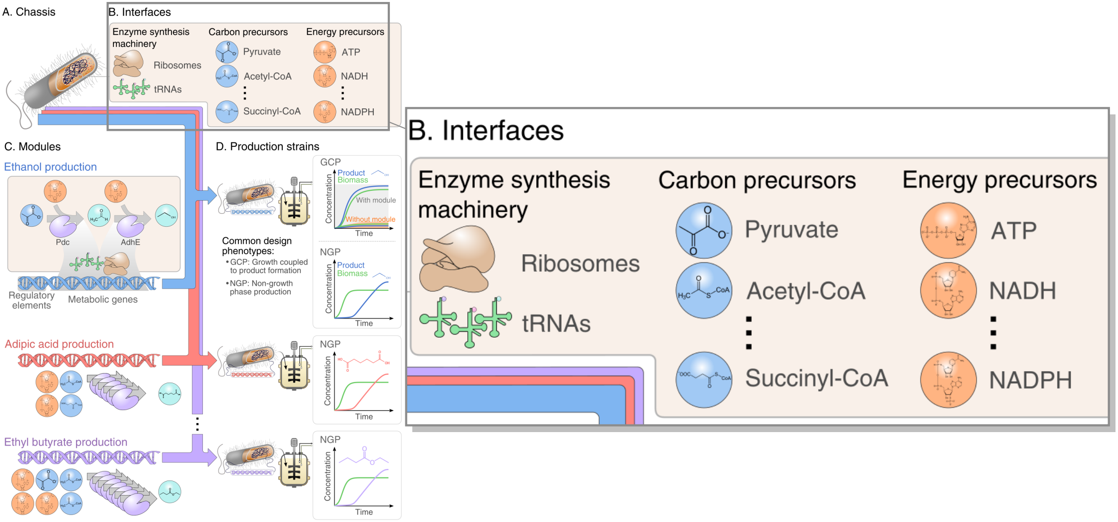
The second wave:

- Basic parts and modules need to be integrated to create systems-level circuitry
- Develop abstract engineering principles and potentially harness biologically-unique features such as adaptation
- Develop better computational models

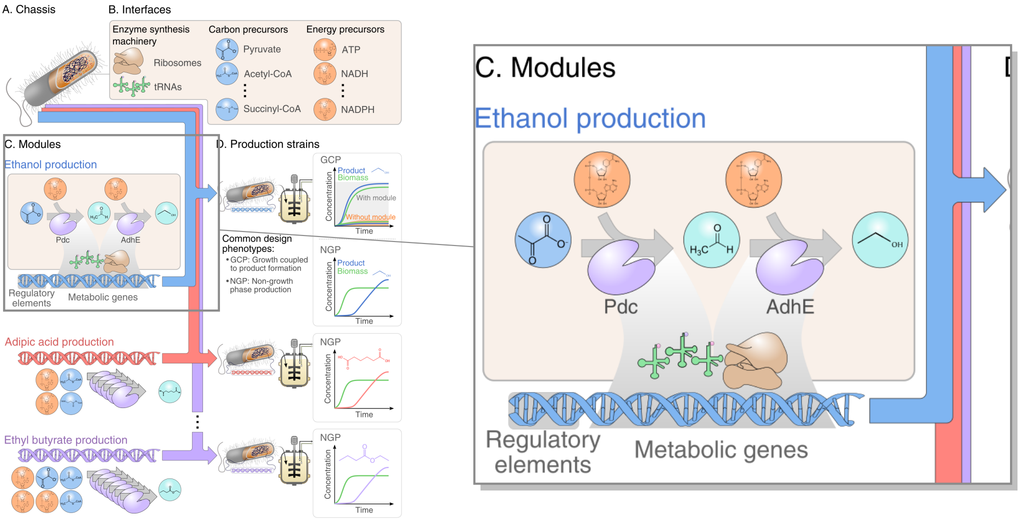
Modular cell (ModCell) biocatalyst



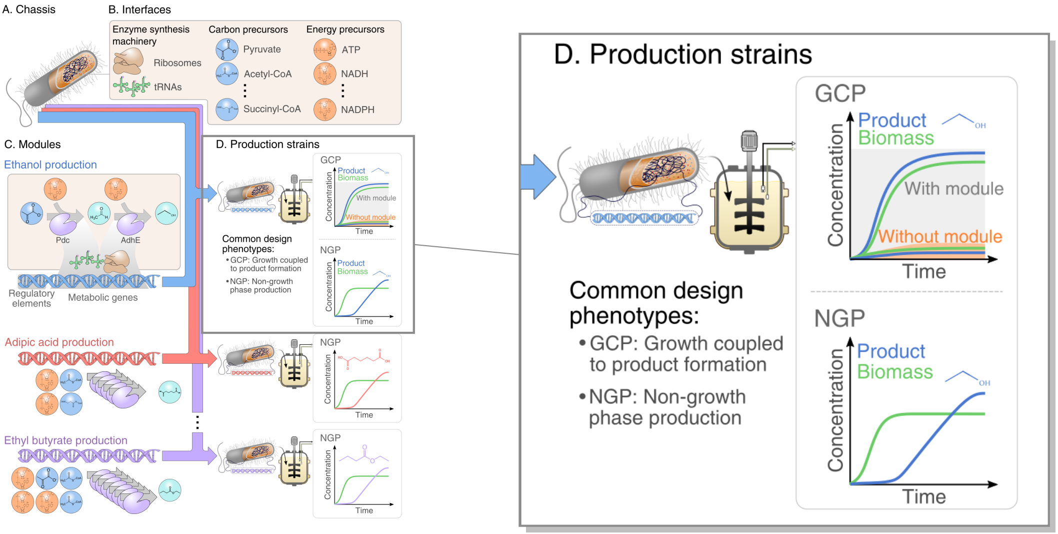
Modular cell (ModCell) biocatalyst



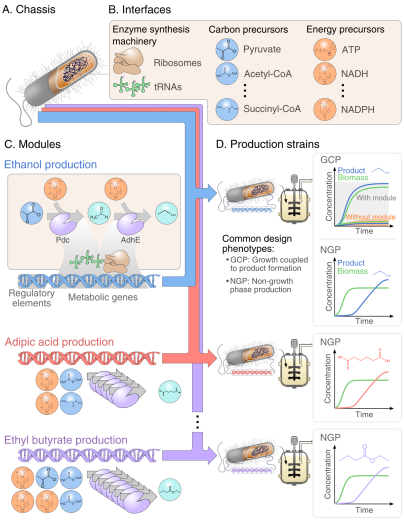
Modular cell (ModCell) biocatalyst



Modular cell (ModCell) biocatalyst



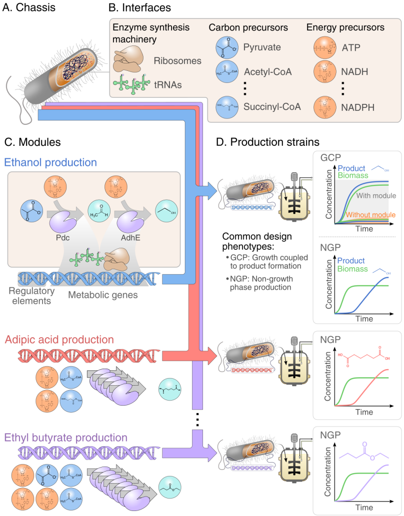
Modular cell (ModCell) biocatalyst



Current approach:

- Integral design of one strain to make a targeted product.

Modular cell (ModCell) biocatalyst



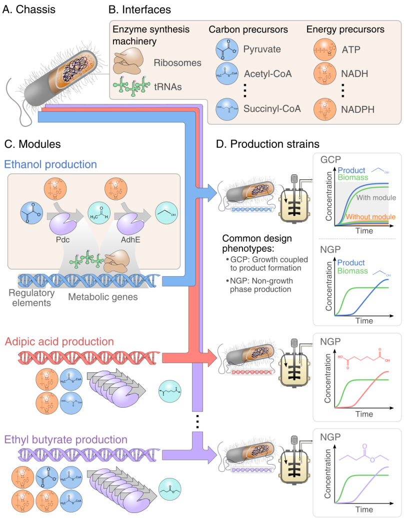
Current approach:

- Integral design of one strain to make a targeted product.

Advantages of modular cell design:

- Efficiency: Combine common elements among the different target phenotypes in the chassis, reducing redundant engineering efforts.

Modular cell (ModCell) biocatalyst



Current approach:

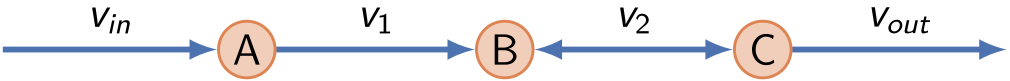
- Integral design of one strain to make a targeted product.

Advantages of modular cell design:

- Efficiency: Combine common elements among the different target phenotypes in the chassis, reducing redundant engineering efforts.
- Predictability: Define and re-uses chassis interfaces to operate with modules, increasing robustness.

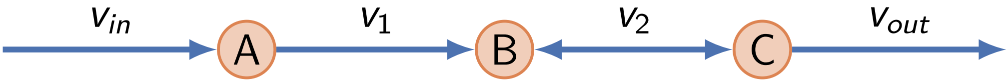
Outline

Basics of constraint-based modeling



Metabolic fluxes v_j (mmol/gCDW/hr)

Basics of constraint-based modeling



Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$\text{A:} \quad \frac{dC_A}{dt} = v_{in} - v_1$$

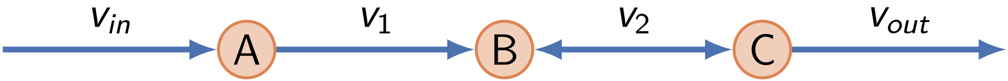
Basics of constraint-based modeling

$$\xrightarrow{v_{in}} A \xrightarrow{v_1} B \xleftrightarrow{v_2} C \xrightarrow{v_{out}}$$

Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$\text{A:} \quad \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0$$

Basics of constraint-based modeling



Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$\begin{aligned}
\text{A:} \quad & \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0 \\
\text{B:} \quad & \frac{dC_B}{dt} = v_1 - v_2 = 0 \\
\text{C:} \quad & \frac{dC_C}{dt} = v_2 - v_{out} = 0
\end{aligned}$$

Basics of constraint-based modeling

$$\xrightarrow{v_{in}} A \xrightarrow{v_1} B \xleftrightarrow{v_2} C \xrightarrow{v_{out}}$$

Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$\begin{aligned}
\text{A:} \quad & \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0 \\
\text{B:} \quad & \frac{dC_B}{dt} = v_1 - v_2 = 0 \\
\text{C:} \quad & \frac{dC_C}{dt} = v_2 - v_{out} = 0
\end{aligned}$$

Lower and upper bounds for each reaction:

Basics of constraint-based modeling

$$\xrightarrow{v_{in}} A \xrightarrow{v_1} B \xleftrightarrow{v_2} C \xrightarrow{v_{out}}$$

Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$
\begin{aligned}
\text{A:} \quad & \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0 \\
\text{B:} \quad & \frac{dC_B}{dt} = v_1 - v_2 = 0 \\
\text{C:} \quad & \frac{dC_C}{dt} = v_2 - v_{out} = 0
\end{aligned}
$$

Lower and upper bounds for each reaction:

$$10 \leq v_{in} \leq 10 \qquad \leftarrow \text{Specify measured flux}$$

Basics of constraint-based modeling

$$\xrightarrow{v_{in}} A \xrightarrow{v_1} B \xleftrightarrow{v_2} C \xrightarrow{v_{out}}$$

Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

$$\text{A:} \quad \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0$$

$$\text{B:} \quad \frac{dC_B}{dt} = v_1 - v_2 = 0$$

$$\text{C:} \quad \frac{dC_C}{dt} = v_2 - v_{out} = 0$$

Lower and upper bounds for each reaction:

$$10 \leq v_{in} \leq 10 \quad \leftarrow \text{Specify measured flux}$$

$$0 \leq v_1 \leq 1000 \quad \leftarrow \text{Irreversible reaction}$$

Basics of constraint-based modeling

$$\xrightarrow{v_{in}} A \xrightarrow{v_1} B \xleftrightarrow{v_2} C \xrightarrow{v_{out}}$$

Metabolic fluxes v_j (mmol/gCDW/hr)
Mass balance for each metabolite in the network:

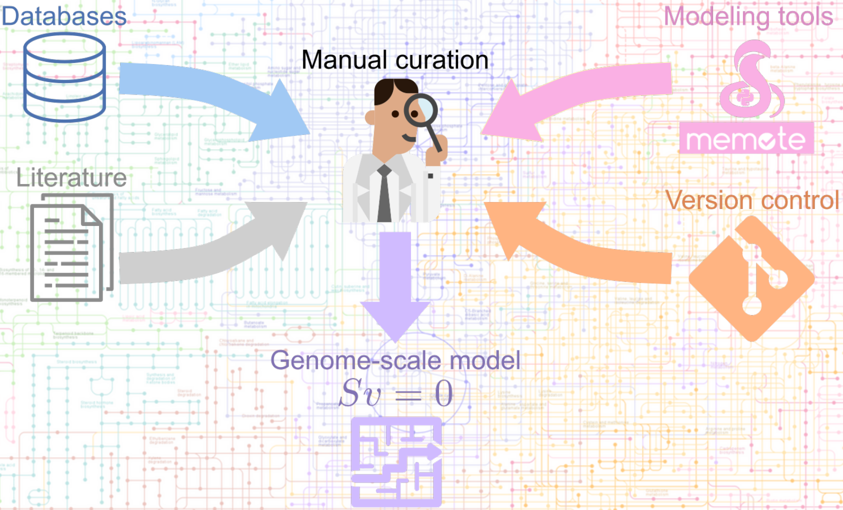
$$\begin{aligned}
\text{A:} \quad & \frac{dC_A}{dt} = v_{in} - v_1 \underset{\text{s.s.}}{=} 0 \\
\text{B:} \quad & \frac{dC_B}{dt} = v_1 - v_2 = 0 \\
\text{C:} \quad & \frac{dC_C}{dt} = v_2 - v_{out} = 0
\end{aligned}$$

Lower and upper bounds for each reaction:

$$\begin{aligned}
10 \leq v_{in} & \leq 10 && \leftarrow \text{Specify measured flux} \\
0 \leq v_1 & \leq 1000 && \leftarrow \text{Irreversible reaction} \\
-1000 \leq v_2 & \leq 1000 && \leftarrow \text{Reversible reaction} \\
0 \leq v_{out} & \leq 1000
\end{aligned}$$

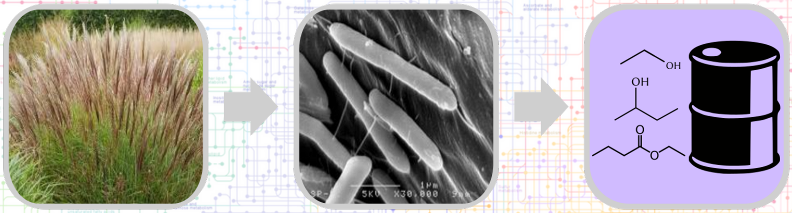
Constraint-based modeling at the genome-scale

With genomic and bibliomic information, constraint-based stoichiometric models can be built at the genome-scale.



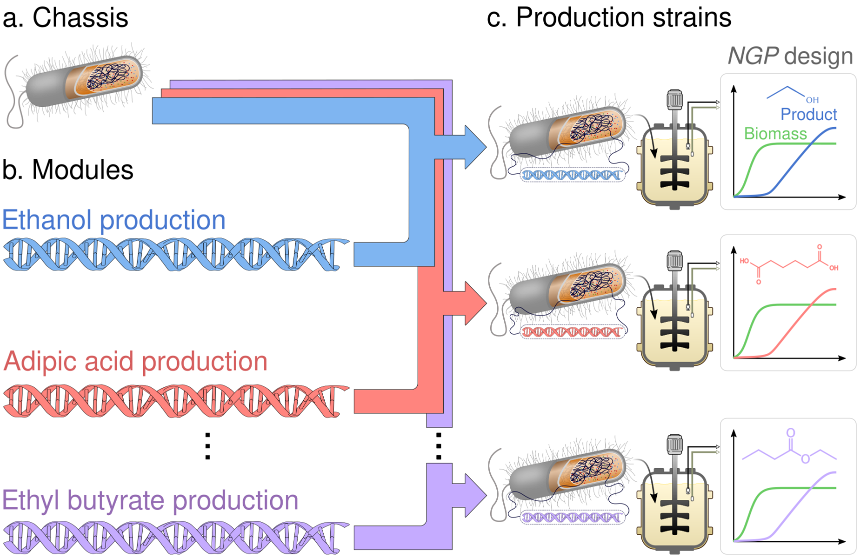
Constraint-based modeling at the genome-scale

Plant biomass — *C. thermocellum* — Biochemicals



Model	iCBI665	iML1515	Yeast 7.6	Recon3D
Organism	*C. thermocellum*	*E. coli*	*S. cerevisiae*	*H. Sapiens*
Genes	665	1515	1149	2248
Metabolites	795	1877	3991	5835
Reactions	854	2712	2691	10600

Modular cell biocatalyst



Mathematical formulation of ModCell

$$
\begin{aligned}
&\max_{y_j, z_{jk}} \quad (f_1, f_2, \ldots, f_{|\mathcal{K}|})^T \quad \text{s.t.} \\
&\quad f_k \in \text{arg max} \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right. \\
&\qquad \sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0 && \text{for all } i \in \mathcal{I}_k \\
&\qquad l_{jk} \leq v_{jk} \leq u_{jk} && \text{for all } j \in \mathcal{J}_k \\
&\qquad l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk} && \text{for all } j \in \mathcal{C} \\
&\qquad \left. \text{where } d_{jk} = y_j \vee z_{jk} \right\} && \text{for all } k \in \mathcal{K} \\
&\quad \sum_{j \in \mathcal{C}} (1 - y_j) \leq \alpha \\
&\quad \sum_{j \in \mathcal{C}} z_{jk} \leq \beta_k && \text{for all } k \in \mathcal{K} \\
&\quad z_{jk} \leq (1 - y_j) && \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}
\end{aligned}
$$

Mathematical formulation of ModCell

$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \dots, f_{|\mathcal{K}|})^T \quad \text{s.t.}$$

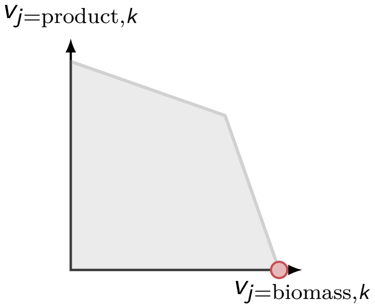
$$\boxed{f_k \in \text{arg max} \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.}$$

$$\left.\begin{array}{ll} \boxed{\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} & \text{for all } i \in \mathcal{I}_k \\ \boxed{l_{jk} \leq v_{jk} \leq u_{jk}} & \text{for all } j \in \mathcal{J}_k \\ l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk} & \text{for all } j \in \mathcal{C} \\ \text{where } d_{jk} = y_j \vee z_{jk} & \end{array}\right\} \quad \text{for all } k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \leq \alpha$$

$$\sum_{j \in \mathcal{C}} z_{jk} \leq \beta_k \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \leq (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$



Mathematical formulation of ModCell

$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \dots, f_{|\mathcal{K}|})^T \quad \text{s.t.}$$

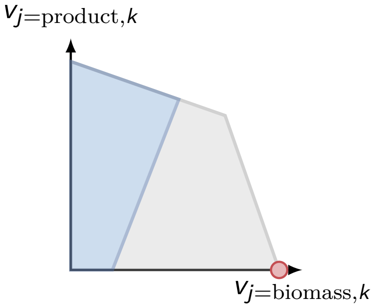
$$\boxed{f_k \in \arg\max \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.}$$

$$\left.\begin{array}{ll} \boxed{\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} & \text{for all } i \in \mathcal{I}_k \\ \boxed{l_{jk} \le v_{jk} \le u_{jk}} & \text{for all } j \in \mathcal{J}_k \\ \boxed{l_{jk} d_{jk} \le v_{jk} \le u_{jk} d_{jk}} & \text{for all } j \in \mathcal{C} \\ \text{where } \boxed{d_{jk} = y_j \vee z_{jk}} & \end{array}\right\} \quad \text{for all } k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \le \alpha$$

$$\sum_{j \in \mathcal{C}} z_{jk} \le \beta_k \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \le (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$



Mathematical formulation of ModCell

$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \dots, f_{|\mathcal{K}|})^T \quad \text{s.t.}$$

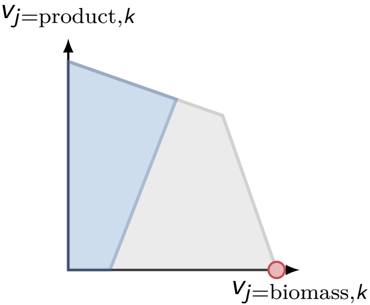
$$\boxed{f_k \in \text{arg max}\left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.}$$

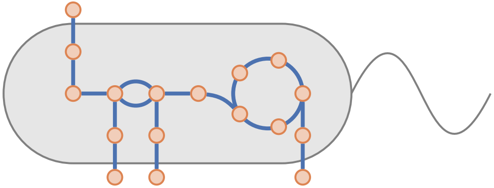
$$\left.\begin{aligned}
&\boxed{\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} && \text{for all } i \in \mathcal{I}_k \\
&\boxed{l_{jk} \le v_{jk} \le u_{jk}} && \text{for all } j \in \mathcal{J}_k \\
&\boxed{l_{jk} d_{jk} \le v_{jk} \le u_{jk} d_{jk}} && \text{for all } j \in \mathcal{C} \\
&\text{where } \boxed{d_{jk} = y_j \vee z_{jk}} &&
\end{aligned}\right\} \quad \text{for all } k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \le \alpha$$

$$\sum_{j \in \mathcal{C}} z_{jk} \le \beta_k \qquad \text{for all } k \in \mathcal{K}$$

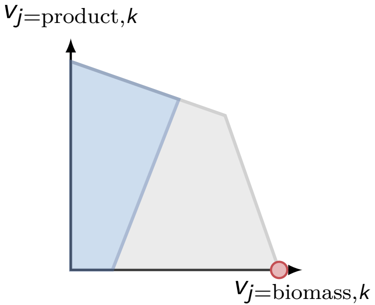
$$z_{jk} \le (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$

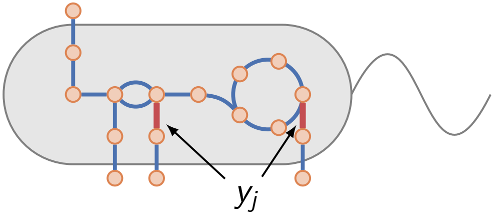




Mathematical formulation of ModCell

$$
\begin{aligned}
&\max_{y_j, z_{jk}} \quad (f_1, f_2, \ldots, f_{|\mathcal{K}|})^T \quad \text{s.t.} \\
&\boxed{f_k \in \arg\max \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.} \\
&\qquad \left.
\begin{array}{ll}
\boxed{\displaystyle\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} & \text{for all } i \in \mathcal{I}_k \\
\boxed{l_{jk} \leq v_{jk} \leq u_{jk}} & \text{for all } j \in \mathcal{J}_k \\
\boxed{l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk}} & \text{for all } j \in \mathcal{C} \\
\text{where } \boxed{d_{jk} = y_j \vee z_{jk}} &
\end{array}
\right\} \quad \text{for all } k \in \mathcal{K} \\
&\sum_{j \in \mathcal{C}} (1 - y_j) \leq \alpha \\
&\sum_{j \in \mathcal{C}} z_{jk} \leq \beta_k \qquad \text{for all } k \in \mathcal{K} \\
&z_{jk} \leq (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}
\end{aligned}
$$





Mathematical formulation of ModCell

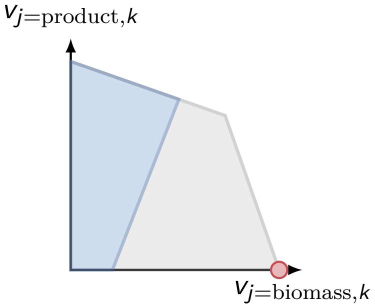
$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \dots, f_{|\mathcal{K}|})^T \quad \text{s.t.}$$

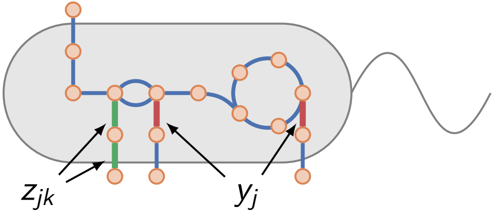
$$\left.\begin{array}{ll}
\boxed{f_k \in \arg\max \left\{ \dfrac{1}{f_k^{max}} \displaystyle\sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.} & \\
\boxed{\displaystyle\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} & \text{for all } i \in \mathcal{I}_k \\
\boxed{l_{jk} \le v_{jk} \le u_{jk}} & \text{for all } j \in \mathcal{J}_k \\
\boxed{l_{jk} d_{jk} \le v_{jk} \le u_{jk} d_{jk}} & \text{for all } j \in \mathcal{C} \\
\text{where } \boxed{d_{jk} = y_j \vee z_{jk}} \quad \Bigg\} & \text{for all } k \in \mathcal{K}
\end{array}\right.$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \le \alpha$$

$$\sum_{j \in \mathcal{C}} z_{jk} \le \beta_k \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \le (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$





Mathematical formulation of ModCell

$$\max_{y_j, z_{jk}} \quad (f_1, f_2, \ldots, f_{|\mathcal{K}|})^T \quad \text{s.t.}$$

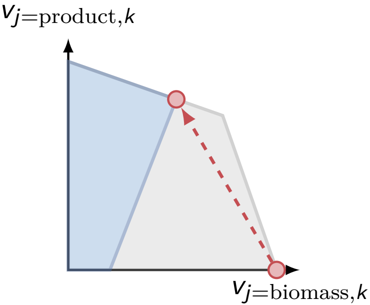
$$f_k \in \arg\max \left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.$$

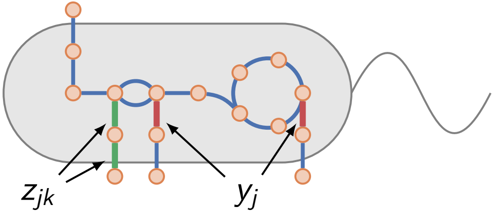
$$\left. \begin{array}{ll} \sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0 & \text{for all } i \in \mathcal{I}_k \\ l_{jk} \leq v_{jk} \leq u_{jk} & \text{for all } j \in \mathcal{J}_k \\ l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk} & \text{for all } j \in \mathcal{C} \\ \text{where } d_{jk} = y_j \vee z_{jk} & \end{array} \right\} \quad \text{for all } k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \leq \alpha$$

$$\sum_{j \in \mathcal{C}} z_{jk} \leq \beta_k \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \leq (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$





Mathematical formulation of ModCell

$$\boxed{\max_{y_j, z_{jk}} \quad (f_1, f_2, \ldots, f_{|\mathcal{K}|})^T \quad \text{s.t.}}$$

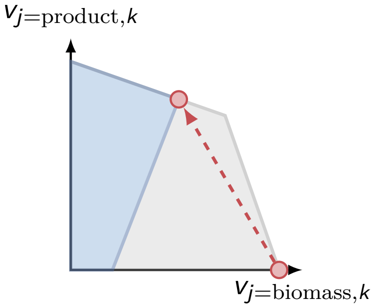
$$\boxed{f_k \in \text{arg max}\left\{ \frac{1}{f_k^{max}} \sum_{j \in \mathcal{J}_k} c_{jk} v_{jk} \quad \text{s.t.} \right.}$$

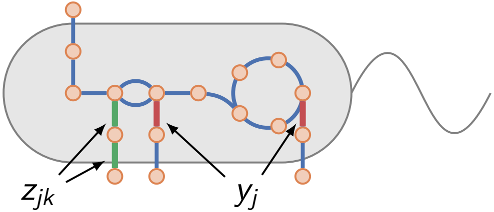
$$\left.\begin{array}{ll}
\boxed{\sum_{j \in \mathcal{J}_k} S_{ijk} v_{jk} = 0} & \text{for all } i \in \mathcal{I}_k \\
\boxed{l_{jk} \leq v_{jk} \leq u_{jk}} & \text{for all } j \in \mathcal{J}_k \\
\boxed{l_{jk} d_{jk} \leq v_{jk} \leq u_{jk} d_{jk}} & \text{for all } j \in \mathcal{C} \\
\text{where } \boxed{d_{jk} = y_j \vee z_{jk}}
\end{array}\right\} \quad \text{for all } k \in \mathcal{K}$$

$$\sum_{j \in \mathcal{C}} (1 - y_j) \leq \boxed{\alpha}$$

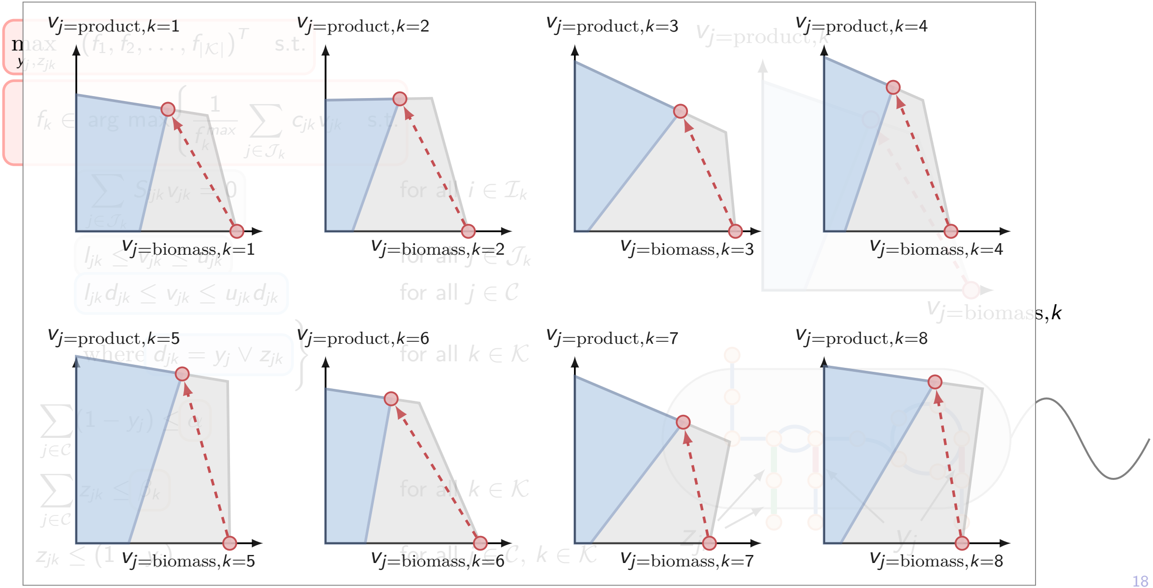
$$\sum_{j \in \mathcal{C}} z_{jk} \leq \boxed{\beta_k} \qquad \text{for all } k \in \mathcal{K}$$

$$z_{jk} \leq (1 - y_j) \qquad \text{for all } j \in \mathcal{C},\ k \in \mathcal{K}$$





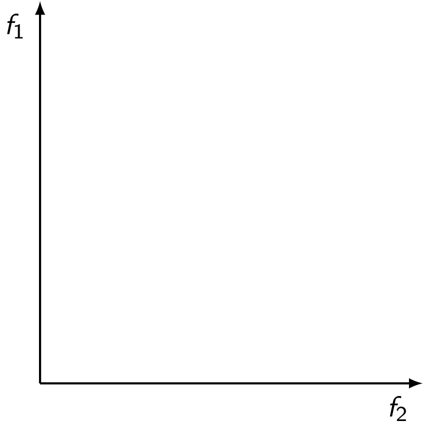
Mathematical formulation of ModCell



What defines a solution?

Multi-objective optimization problem

$$\max_{x \in X} \quad F(x) = (f_1(x), f_2(x), \ldots)^T$$

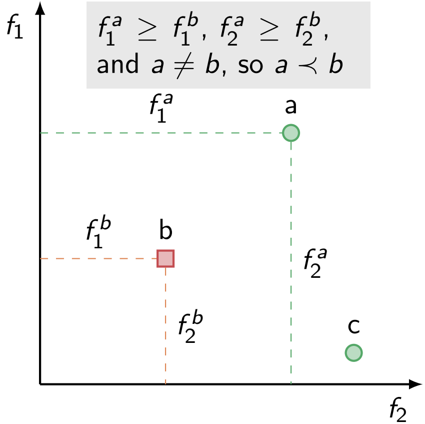


What defines a solution?

Multi-objective optimization problem

$$\max_{x \in X} \quad F(x) = (f_1(x), f_2(x), \ldots)^T$$

$f_1^a \geq f_1^b$, $f_2^a \geq f_2^b$,
and $a \neq b$, so $a \prec b$



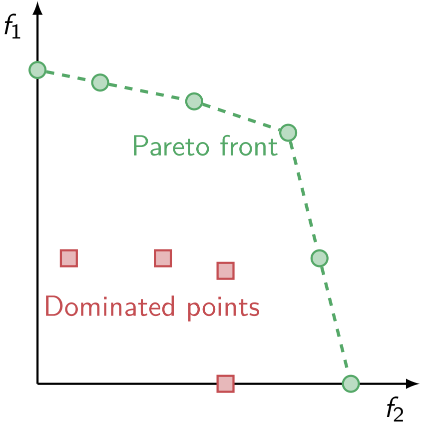
Definition of domination

A vector a dominates another vector b (denoted $a \prec b$) iff $a_i \geq b_i \; \forall i \in \{1, 2, \ldots, K\}$ and $a_i \neq b_i$ for at least one i.

What defines a solution?

Multi-objective optimization problem

$$\max_{x \in X} \quad F(x) = (f_1(x), f_2(x), \ldots)^T$$



Definition of domination

A vector a dominates another vector b (denoted $a \prec b$) iff $a_i \geq b_i \ \forall i \in \{1, 2, \ldots, K\}$ and $a_i \neq b_i$ for at least one i.

Pareto set

$$PS := \{x \in X : \nexists x' \in X, F(x') \prec F(x)\}$$

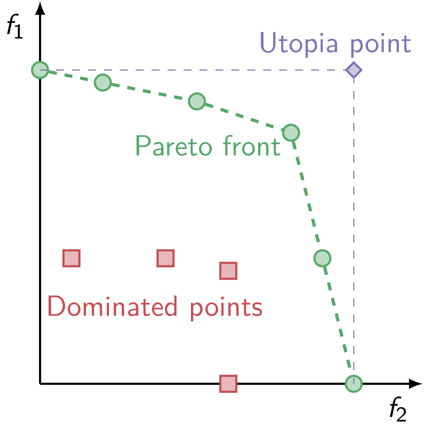
Pareto front

$$PF := \{F(x) : x \in PS\}$$

What defines a solution?

Multi-objective optimization problem

$$\max_{x \in X} \quad F(x) = (f_1(x), f_2(x), \ldots)^T$$



Definition of domination

A vector a dominates another vector b (denoted $a \prec b$) iff $a_i \geq b_i \; \forall i \in \{1, 2, \ldots, K\}$ and $a_i \neq b_i$ for at least one i.

Pareto set

$$PS := \{x \in X : \nexists x' \in X, F(x') \prec F(x)\}$$

Pareto front

$$PF := \{F(x) : x \in PS\}$$

Utopia point

$$max(f_1(x)) = f_1(x) = f_2(x) = ...$$

Solution algorithms for multi-objective optimization

Solution algorithms for multi-objective optimization

MILP (Mixed integer linear programming)

- Convert to single-objective problem
- Need a priori specification of preference
- Optimality guaranteed

Goal attaintment formulation: define a performance target for each objective g_k

$$
\begin{aligned}
\min \quad & \sum_{k \in \mathcal{K}} \delta_k && (1) \\
\text{s.t.} \quad & && \\
& f'_k + \delta_k \geq g_k \ \forall k \in \mathcal{K} && (2) \\
& \delta_k \geq 0 \qquad \forall k \in \mathcal{K} && (3) \\
& f' \in \Omega && (4)
\end{aligned}
$$

Solution algorithms for multi-objective optimization

MILP (Mixed integer linear programming)

- Convert to single-objective problem
- Need a priori specification of preference
- Optimality guaranteed

Goal attaintment formulation: define a performance target for each objective g_k

$$\min \sum_{k \in \mathcal{K}} \delta_k \tag{1}$$

s.t.

$$f'_k + \delta_k \geq g_k \quad \forall k \in \mathcal{K} \tag{2}$$

$$\delta_k \geq 0 \qquad \forall k \in \mathcal{K} \tag{3}$$

$$f' \in \Omega \tag{4}$$

MOEA (Multi-objective evolutionary algorithm)

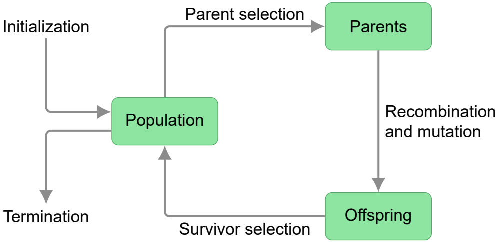
- Directly obtain Pareto front
- Can be easily adapted to different models and design objectives
- Scalable to HPC for problems with many objectives

Basics of genetic algorithms

Population-based heuristic optimization (e.g., Genetic algorithms):

- Individual: Encodes the variables of the problem and hence has an objective value associated with it.
- Operators: Heuristic that modify individuals to enhance their objective values.

The population of individuals is modified with operators to identify potentially optimal solutions to the optimization problem.



Basics of genetic algorithms

Population-based heuristic optimization (e.g., Genetic algorithms):

- Individual: Encodes the variables of the problem and hence has an objective value associated with it.
- Operators: Heuristic that modify individuals to enhance their objective values.

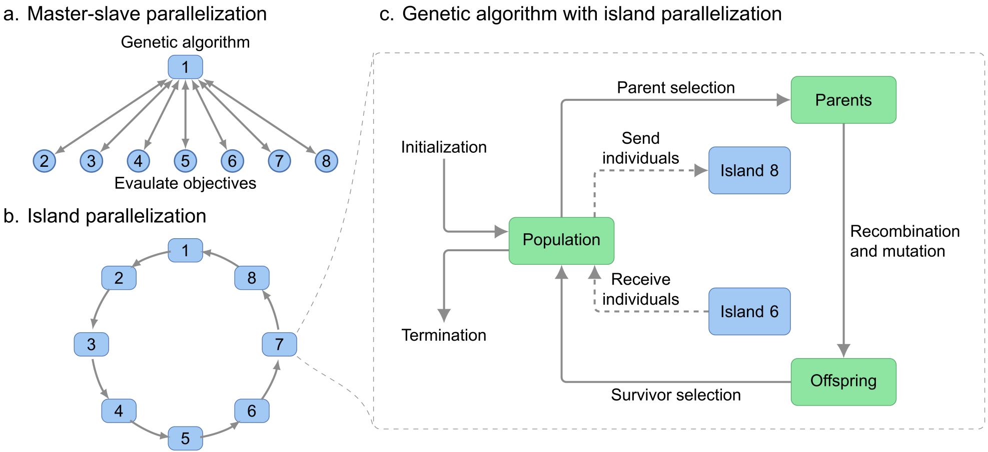
The population of individuals is modified with operators to identify potentially optimal solutions to the optimization problem.

The 2006 NASA ST5 spacecraft antenna. This complicated shape was found by an evolutionary computer design program to create the best radiation pattern.

Scaling MOEA for many-objective problems

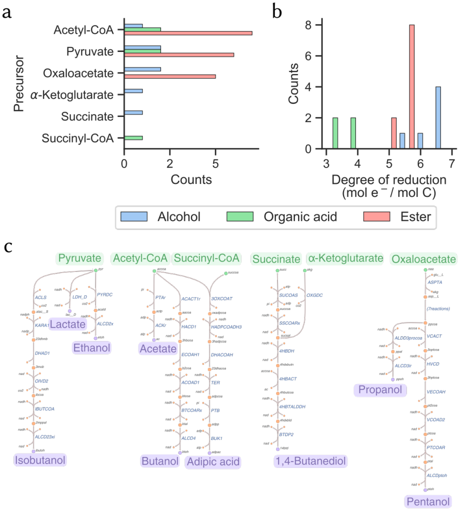
- We wish to solve modular cell design problems with 100s of products
- Many-objective problems are notoriusly difficult to solve and HPC approaches are not well explored in the field
- Develop and benchmark an HPC MOEA

MOEA parallelization

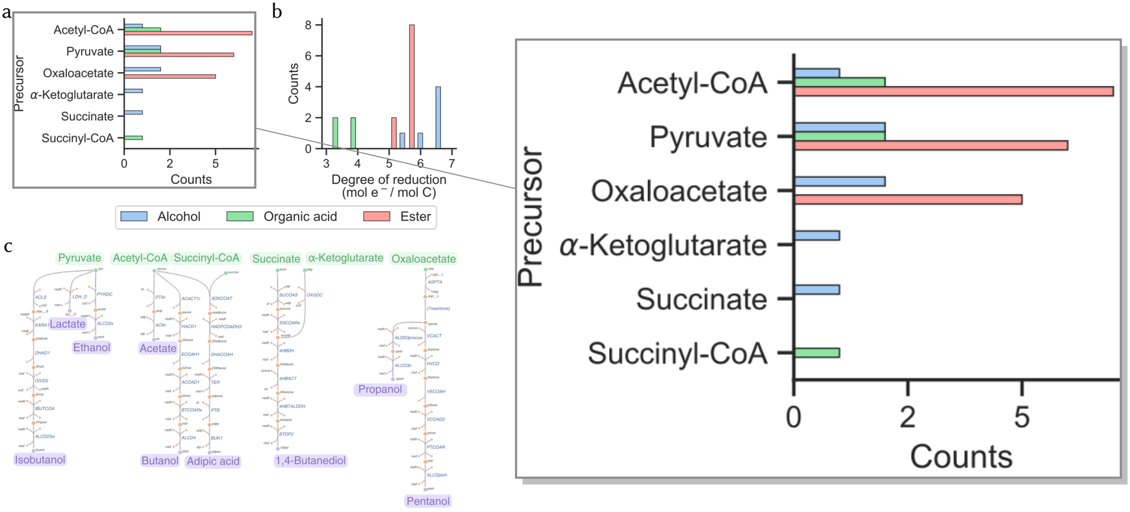


Outline

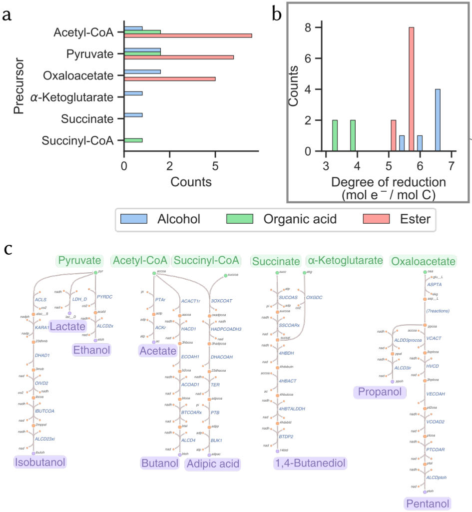
Input: 20 diverse products

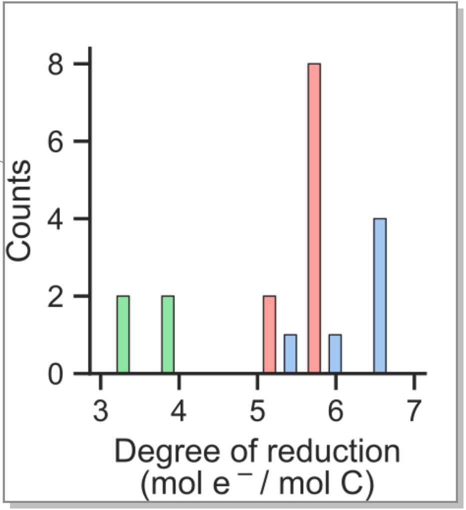


Input: 20 diverse products

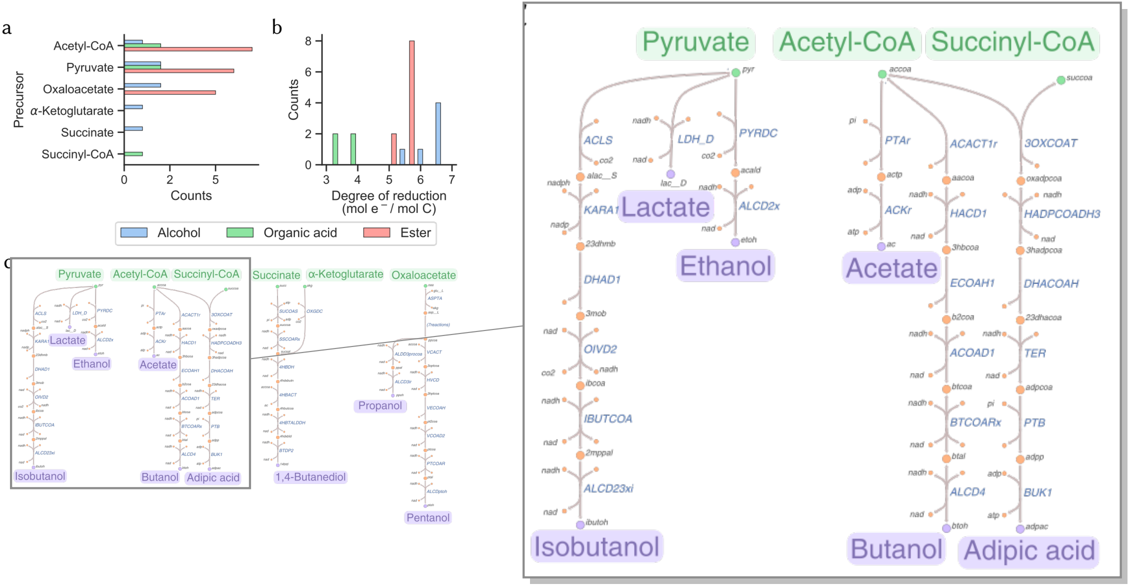


Input: 20 diverse products

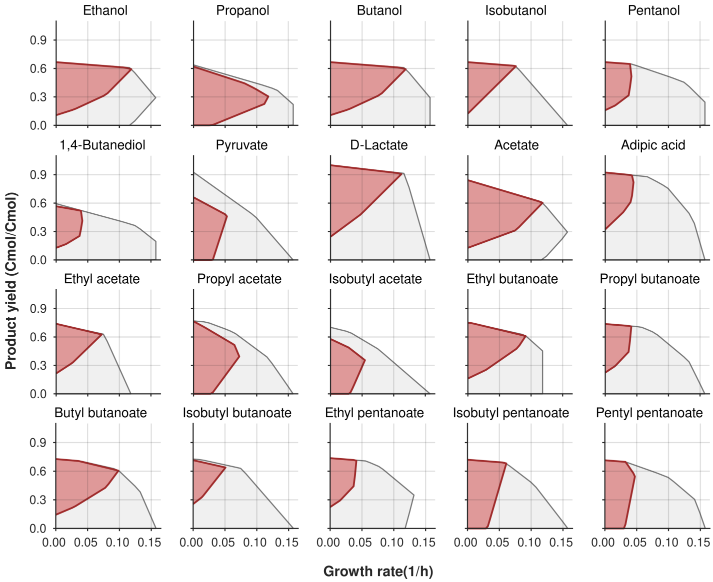




Input: 20 diverse products



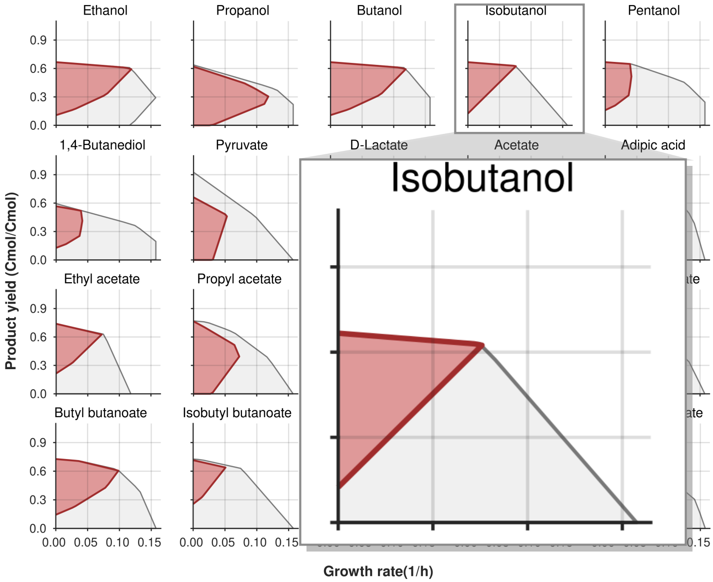
Universal design for 20 products



Phenotypic spaces:

- Represent feasible metabolic states according to stoichiometric constraints
- Gray region: Wild type + production module
- Red region: Designed chassis + production module

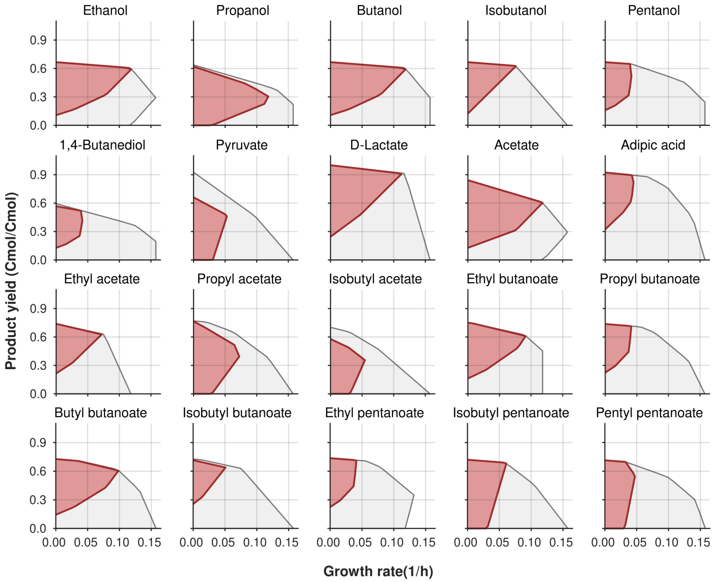
Universal design for 20 products



Phenotypic spaces:

- Represent feasible metabolic states according to stoichiometric constraints
- Gray region: Wild type + production module
- Red region: Designed chassis + production module

Universal design for 20 products

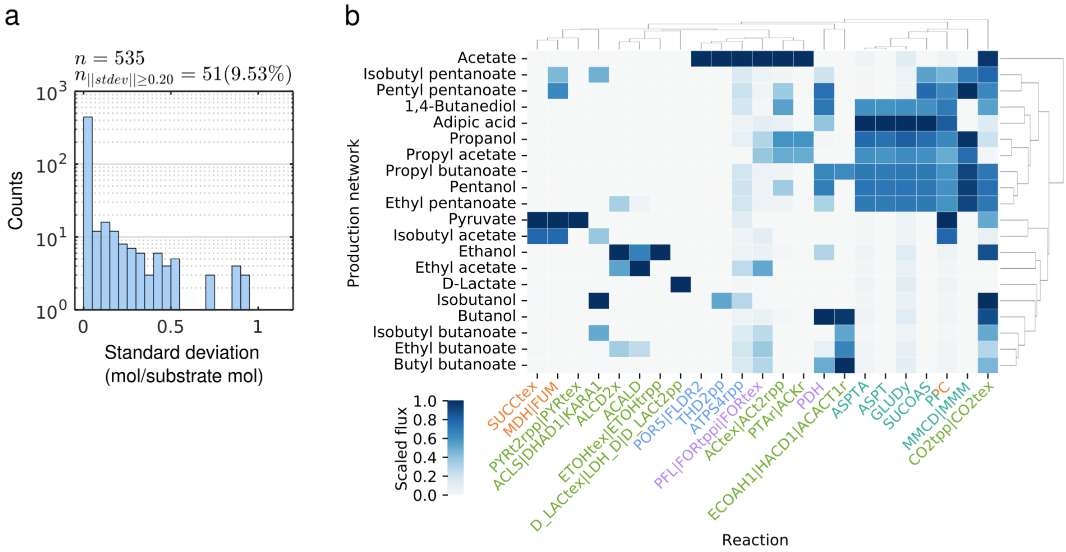


Phenotypic spaces:

- Represent feasible metabolic states according to stoichiometric constraints
- Gray region: Wild type + production module
- Red region: Designed chassis + production module

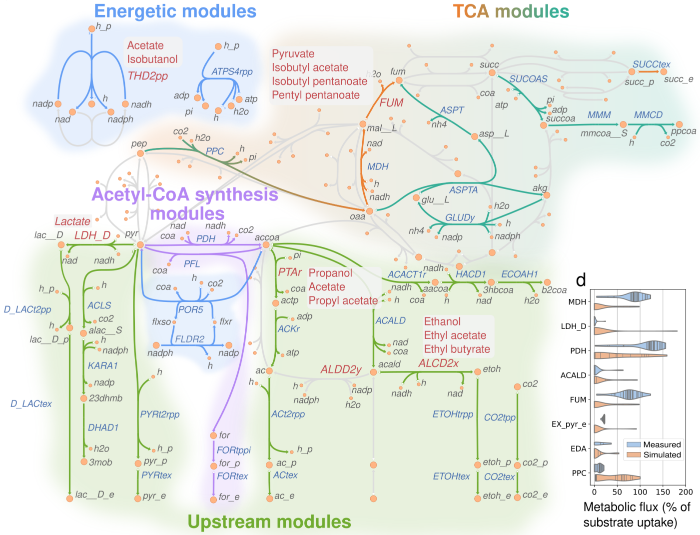
The universal design leads to high product yields at the maximum growth rate for all combinations of chassis and production modules.

Identification of chassis metabolic interfaces

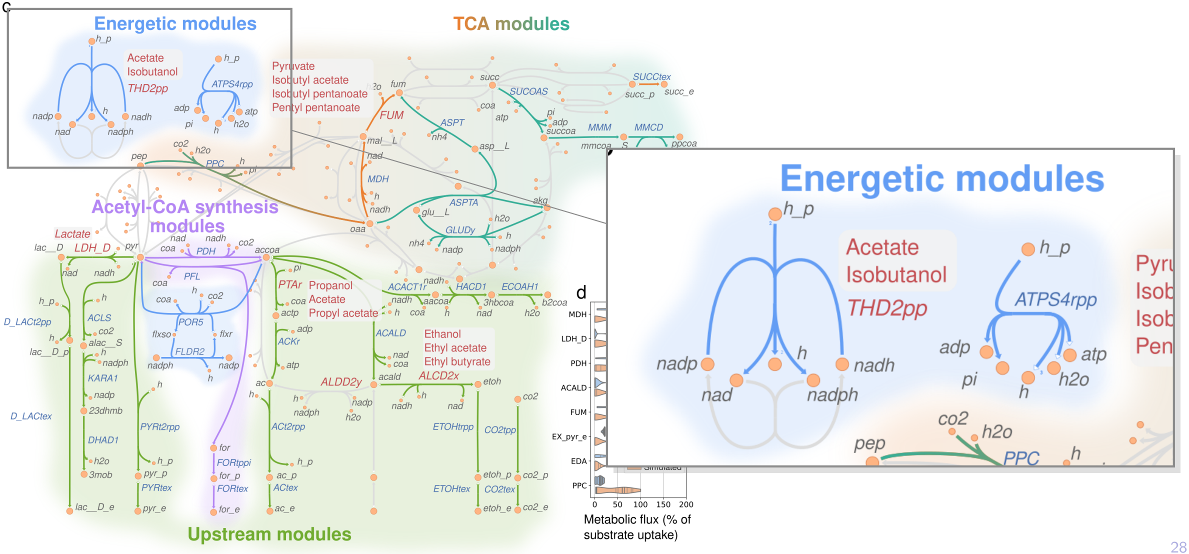


Natural modularity and flexibility of core metabolism enables universal design

c

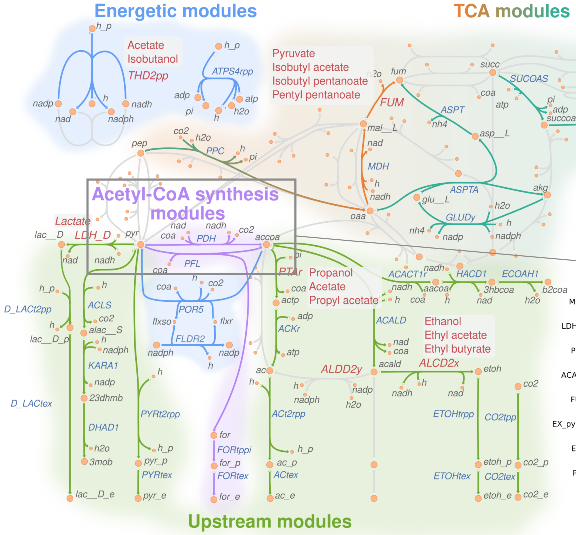


Natural modularity and flexibility of core metabolism enables universal design

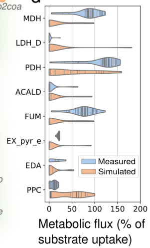


Natural modularity and flexibility of core metabolism enables universal design

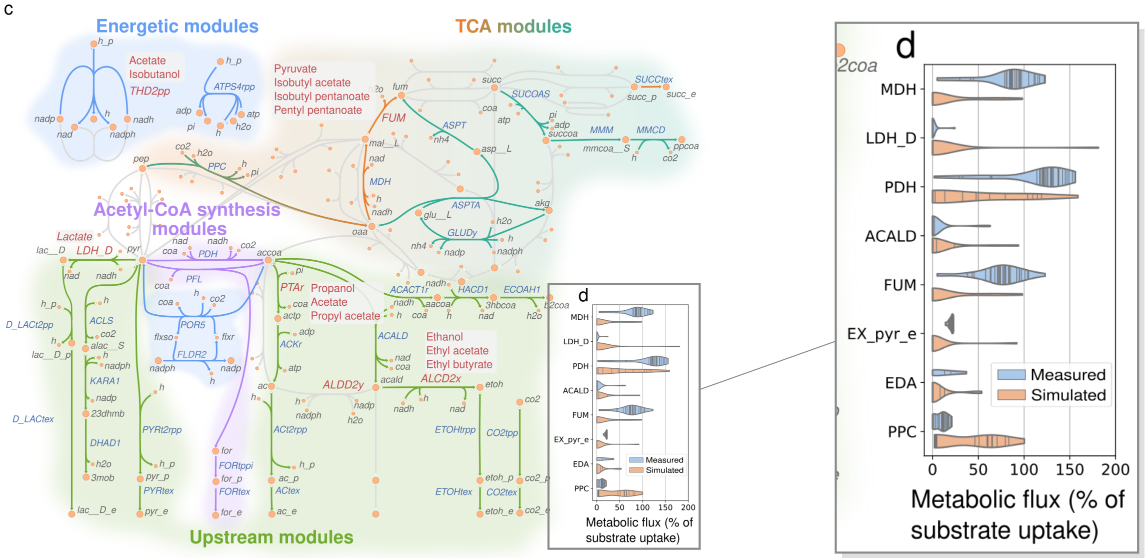
c



d



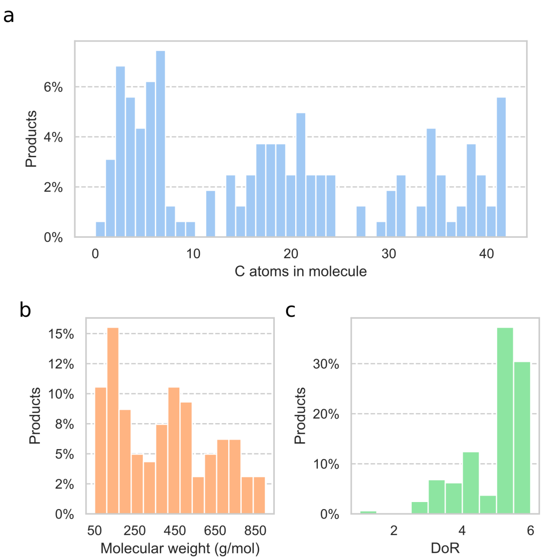
Natural modularity and flexibility of core metabolism enables universal design



Input: 161 endogenous products

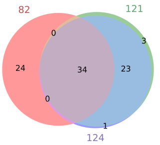
Endogenous metabolites in *E. coli* that are organic and can be coupled to growth under anaerobic conditions.

- Diverse molecule size
- Highly reduced due to anaerobic conditions.

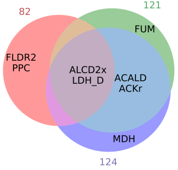


Smallest set of designs to cover product library

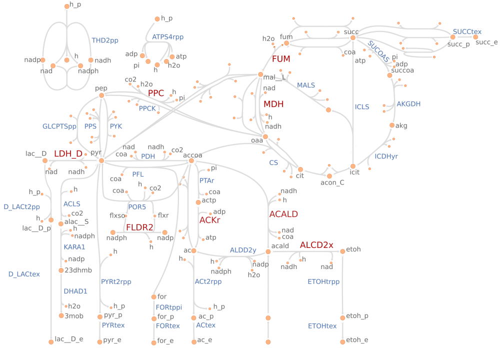
a. Compatible modules



b. Reaction deletions

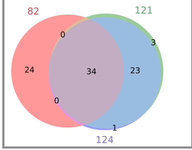


c. Metabolic location of reaction deletions

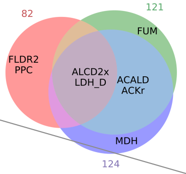


Smallest set of designs to cover product library

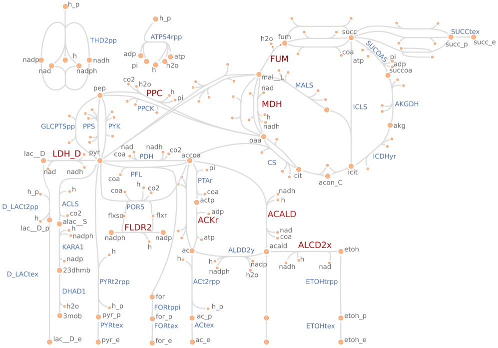
a. Compatible modules



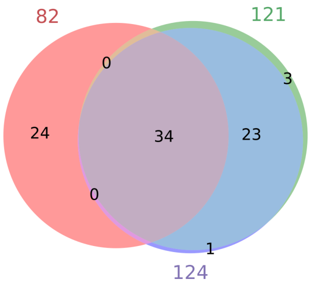
b. Reaction deletions



c. Metabolic location of reaction deletions

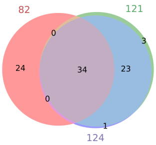


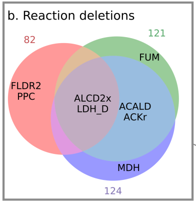
a. Compatible modules



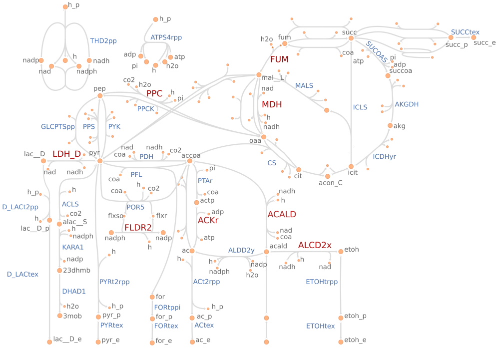
Smallest set of designs to cover product library

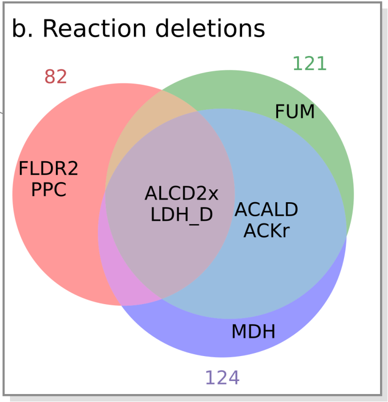
a. Compatible modules





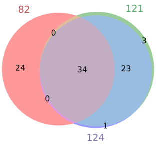
c. Metabolic location of reaction deletions



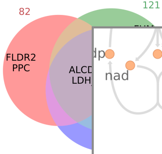


Smallest set of designs to cover product library

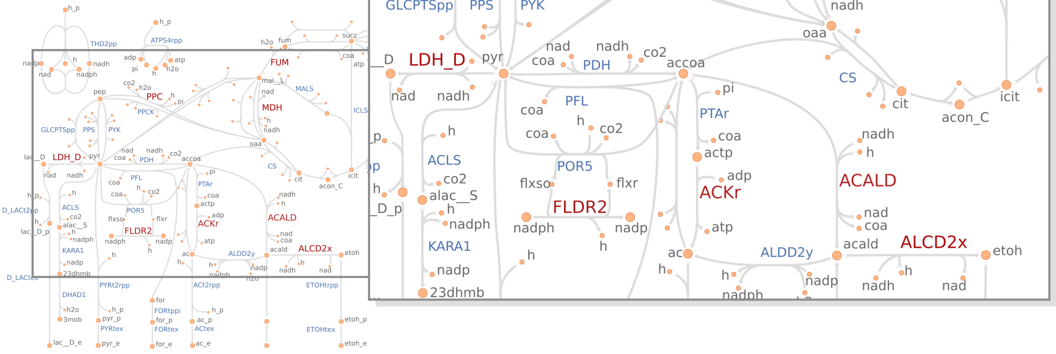
a. Compatible modules



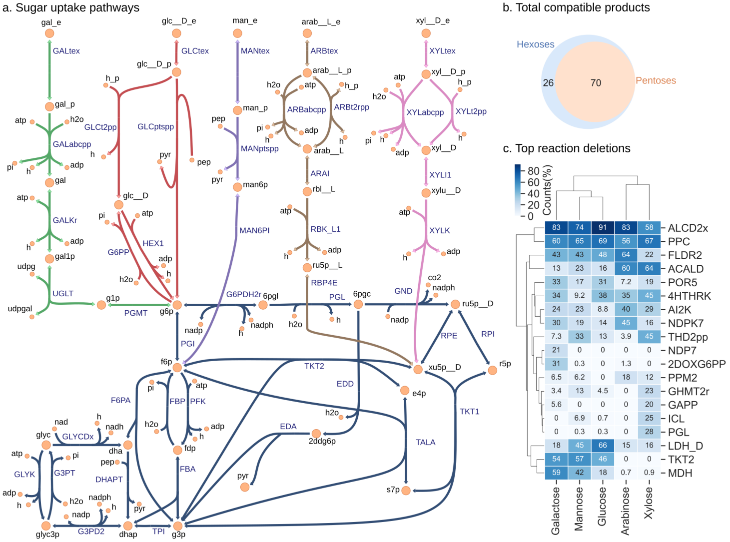
b. Reaction deletions



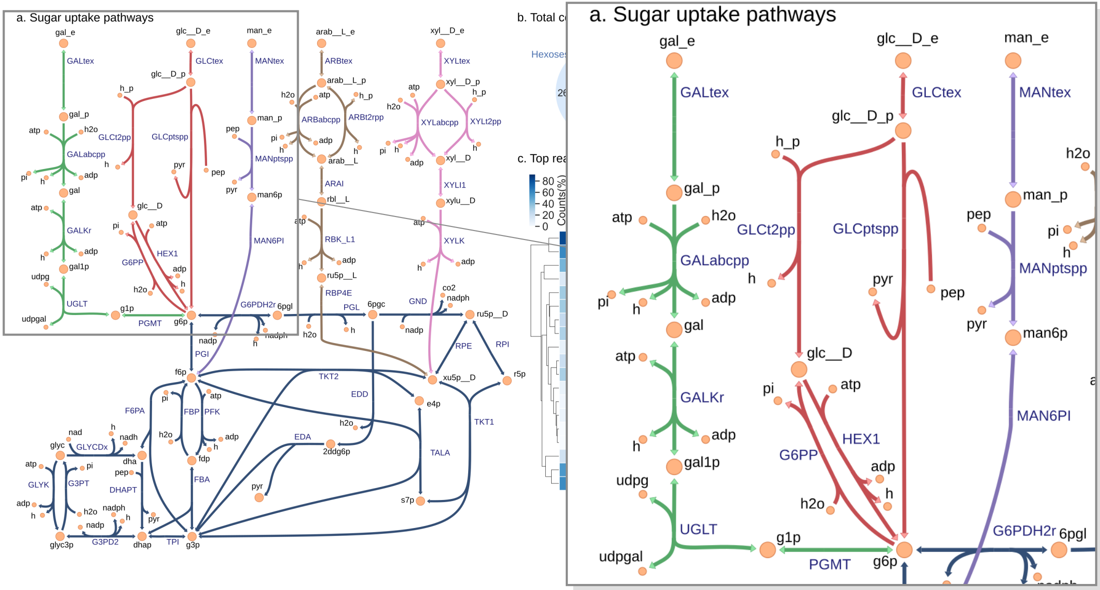
c. Metabolic location of reaction deletions



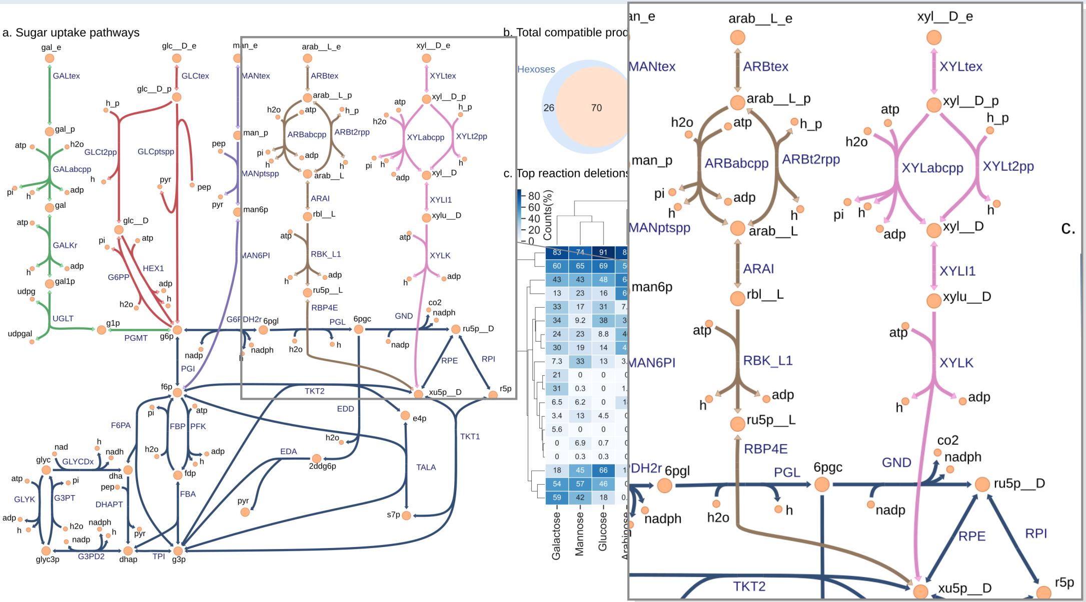
Designs for different carbon sources



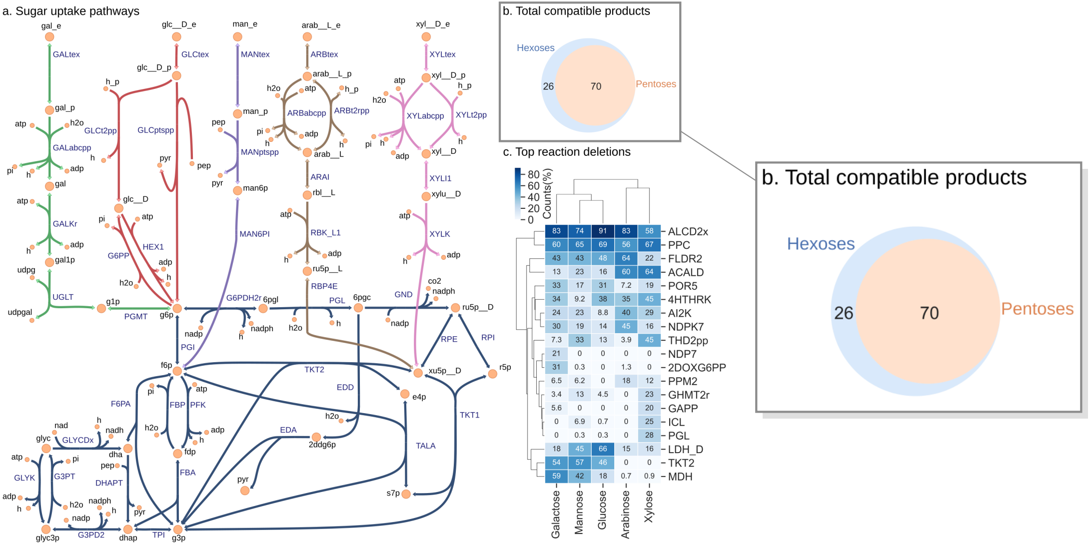
Designs for different carbon sources



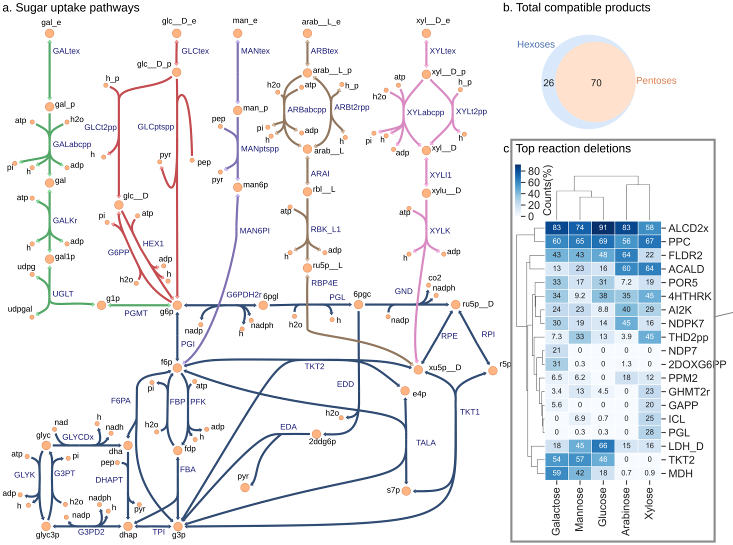
Designs for different carbon sources

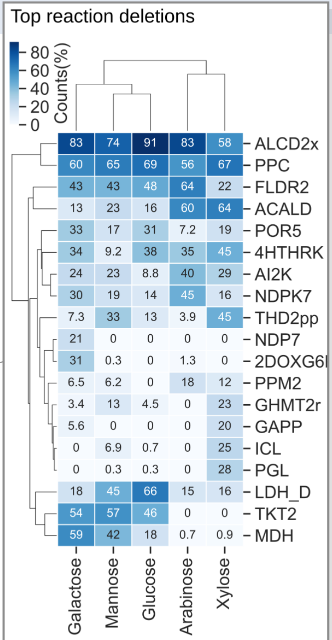


Designs for different carbon sources



Designs for different carbon sources





Outline

ModCell tools

Primer: Metabolic engineering strategies repeat across target molecules

MODCELL: Design strains for individual products and find common manipulations

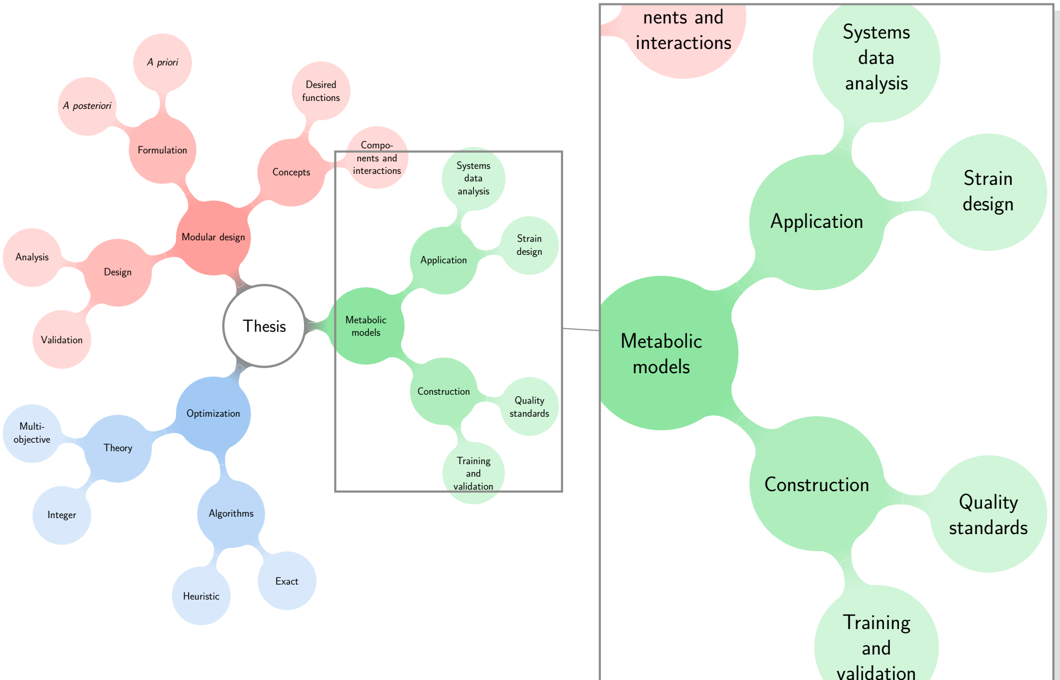
ModCell2: Multi-objective optimization formulation to design Pareto optimal chassis and modules

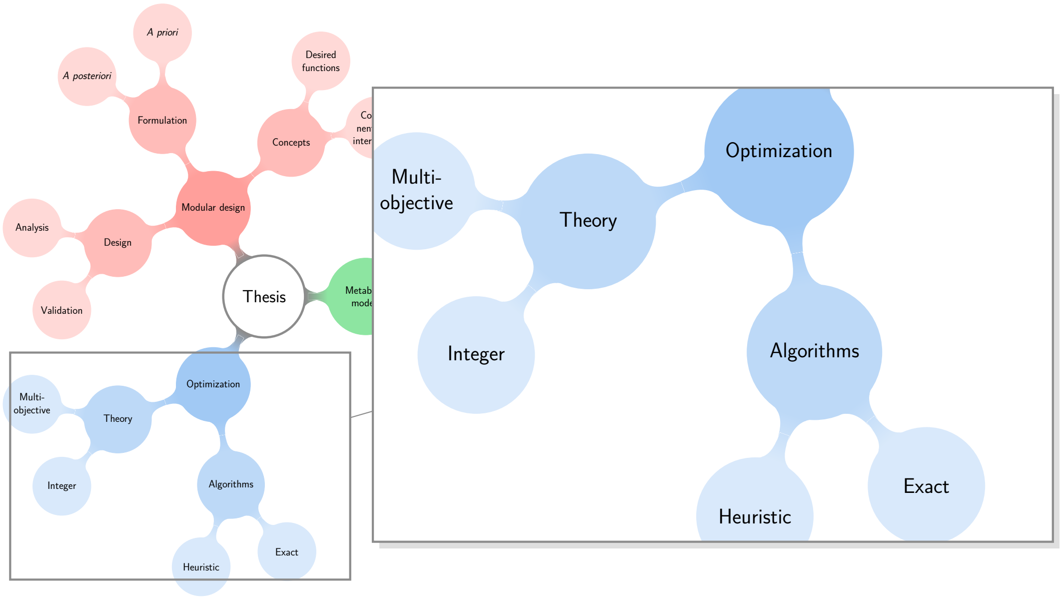
ModCell2-MILP: Identify optimal solutions to design universal chassis and define key metabolic interfaces

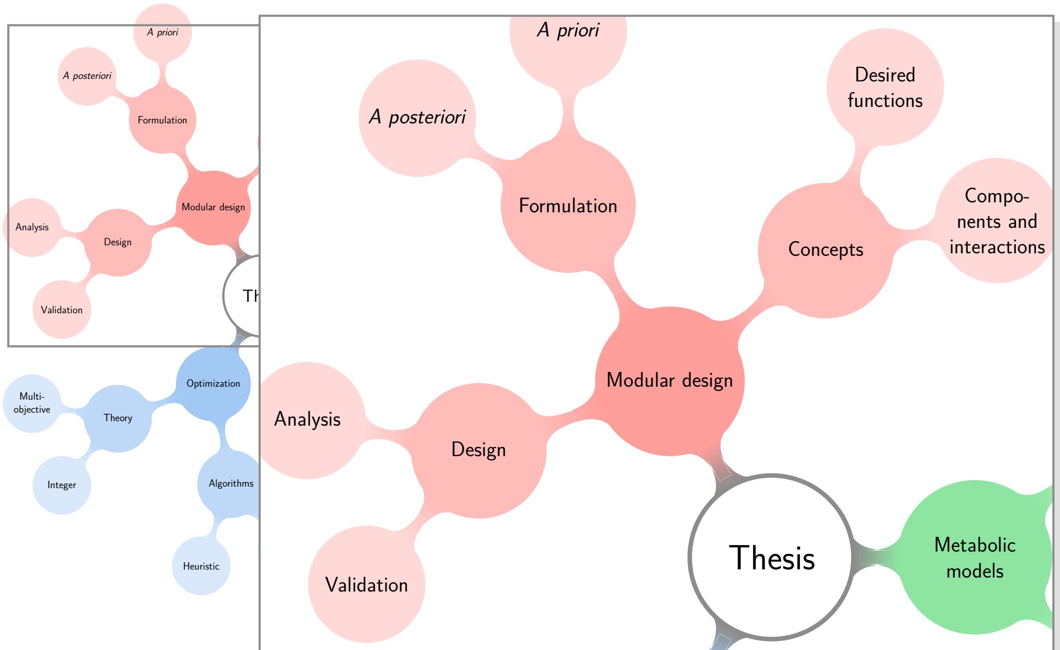
Modcell2-HPC: Design chassis for hundredths of endogenous products providing general principles for synthetic and natural modular design

Modular design principles in metabolic engineering for efficient and robust systems


Thesis
Modular design
Formulation
A priori
A posteriori
Concepts
Desired functions
Compo- nents and interactions
Design
Analysis
Validation
Metabolic models
Application
Systems data analysis
Strain design
Construction
Quality standards
Training and validation
Optimization
Theory
Multi- objective
Integer
Algorithms
Heuristic
Exact


Thesis
Modular design
Formulation
A priori
A posteriori
Concepts
Desired functions
Compo-nents and interactions
Design
Analysis
Validation
Optimization
Theory
Multi-objective
Integer
Algorithms
Heuristic
Exact
Metabolic models
Application
Systems data analysis
Strain design
Construction
Quality standards
Training and validation


Thesis
Modular design
Formulation
A priori
A posteriori
Concepts
Desired functions
Design
Analysis
Validation
Optimization
Theory
Multi-objective
Integer
Algorithms
Heuristic
Exact


A priori
A posteriori
Formulation
Modular design
Analysis
Design
Validation
Th
Optimization
Multi-objective
Theory
Integer
Algorithms
Heuristic
A priori
A posteriori
Desired functions
Formulation
Compo-nents and interactions
Concepts
Modular design
Analysis
Design
Validation
Thesis
Metabolic models

Acknowledgements

Funding Sources

NSF

NSF CAREER

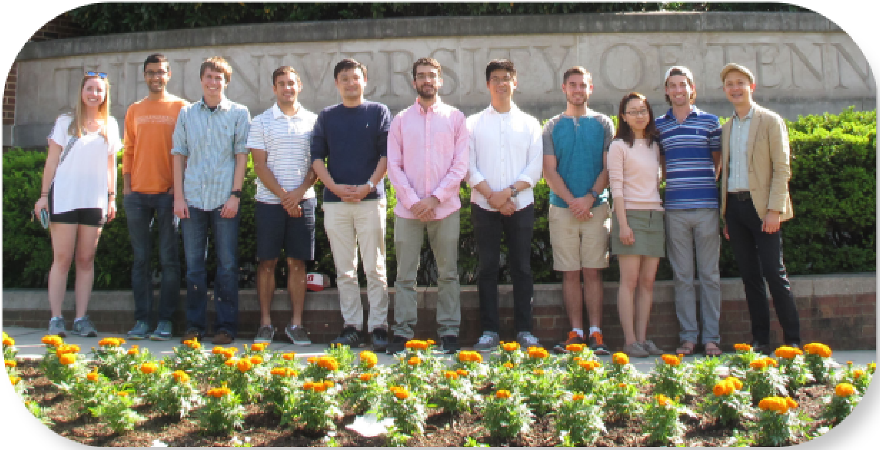
cbi

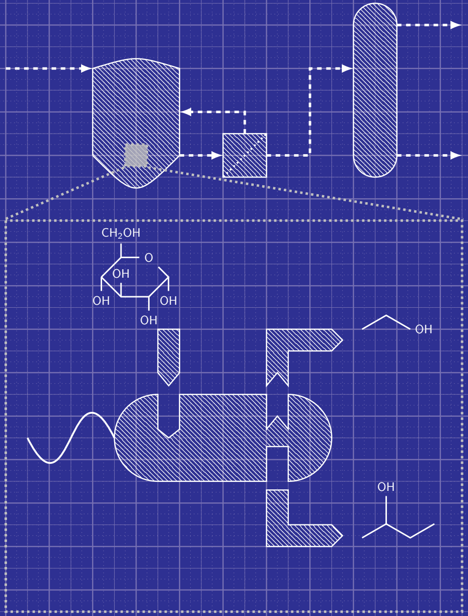
THE CENTER FOR
BIOENERGY
INNOVATION

T

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

Trinh Lab





SG