ModCell2: A Multiobjective Optimization Platform for Modular Strain Design

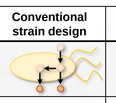
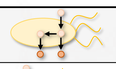
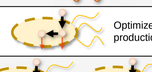
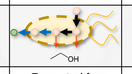
Sergio Garcia[1,2] **and Cong T. Trinh**[1,2,3]

[1] *Department of Chemical and Biomolecular Engineering, University of Tennessee; Knoxville, Tennessee;* [2]*BioEnergy Science Center, Oak Ridge, Tennessee;* [3]*Bredesen Center for Interdisciplinary Research and Education.*

Abstract: Metabolic engineering has recently enabled the use of microbes for industrial production of diverse biochemicals. However, developing an optimal strain for synthesis of one product with the existing technologies is laborious and expensive. To accelerate the process and reduce the cost of strain engineering, we propose the modular cell (ModCell) design principle, which exploits the modular organization of biological networks to design plug-and-play cellular biocatalysts. In this work, we introduce the ModCell2 platform, which uses multiobjective optimization principles and genome scale models for modular cell design. We use ModCell2 to design the biomass-degrading *Clostridium thermocellum* modular cell for growth coupled synthesis of multiple biochemicals, e.g., alcohols and esters from lignocellulosic biomass. We envision that ModCell2 will provide a useful approach for modular cell engineering.

Theory Development

Proposed Modular Strain Design Paradigm

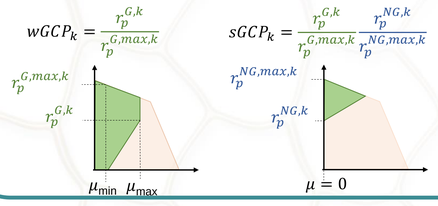
Features	Conventional strain design	Modular strain design
Strain: Parent		
Strain: Chassis	Absent	 Optimized common production phenotypes
Strain: Production		 Module1 Module 2 Module 3
Design-build-test cycle	Repeated for every new product	One time for many products
Mathematical formulation	Single-objective programming	Multiple-objective programming
Design objectives	• Weak growth coupled to product formation (wGCP) • Strong growth coupled to product formation (sGCP) • Non-growth production (NGP)	
Design variables	• Removal or decrease native reaction flux • Increase flux through native or heterologous reactions	

The chassis strain features optimized phenotypes enabling high yield, titer and productivity for products with similar biochemical properties (e.g. precursors, redox state, toxicity mechanisms), allowing for high-throughput assembly of cellular biocatalysts.

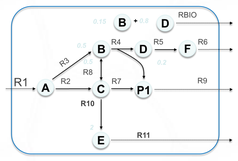
Design Objective (f_k)

- Weak growth coupled to product formation (wGCP):
  - Maximize: Minimum product yield at the maximum growth rate.
- Strong growth coupled to product formation (sGCP):
  - In addition to wGCP, minimum product yield during non-growth phase is also maximized.
- Both objectives allow for adaptative laboratory evolution to improve product yield (experimentally demonstrated).

$$wGCP_k = \frac{r_p^{G,k}}{r_p^{G,max,k}} \qquad sGCP_k = \frac{r_p^{G,k}}{r_p^{G,max,k}} \frac{r_p^{NG,k}}{r_p^{NG,max,k}}$$



Constraint Based Modeling of Metabolism



S_{ij} ≔ Stoichiometric coefficient of metabolite i in reaction j
r_j ≔ Flux of reaction j (mmol/gCDW/hr)
C_i ≔ Concentration of metabolite i
l_j ≔ Lower flux bound of reaction j
u_j ≔ Upper flux bound of reaction j

In – out = acc.

$$\sum_{j \in J} S_{ij} r_j = \frac{dC_i}{dt} = 0 \text{ for all } i \in Metabolites \quad \text{Mass balance}$$

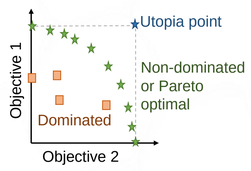
$$l_j \le r_j \le u_j \text{ for all } j \in Reactions \quad \text{Reaction rate bounds}$$

While kinetic models require to estimate or measure kinetic parameters and only return one flux distribution; constraint based models (CBM) only require stoichiometric information, and consider all feasible flux distributions allowed by the mass balance and thermodynamic constraints. The CBM approach has been prominent in systems biology since the early two-thousands.

Multiobjective Optimization Formulation

Maximize $(f_1, f_2, \ldots, f_{|Products|})^T$
Over reaction deletions and module reaction
subject to:
- At most α reaction deletions in chassis

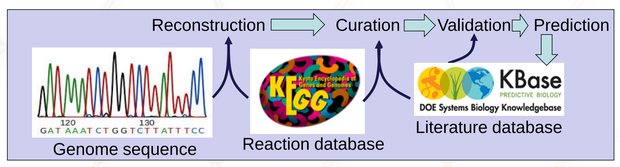
For each production network k:
- At most β_k module reactions
- Metabolite mass balance
- Reaction directionality and flux bounds
- Chassis deletions and module reactions



A. The modular cell design problem is expressed into the interdisciplinary mathematical framework of Multiobjective optimization.
B. The solutions to the modular cell design problem are termed Pareto optimal or efficient, because they require to diminish one objective to increase another.

Case Study: Biofuels in *C. thermocellum*

Clostridium thermocellum DSM1313 Genome Scale Model *iAT601*



The refined genome scale model consists of 871 reactions encoded by 601 genes, spanning central and peripheric metabolic pathways, as well as biomass and cellulosome synthesis pathways.

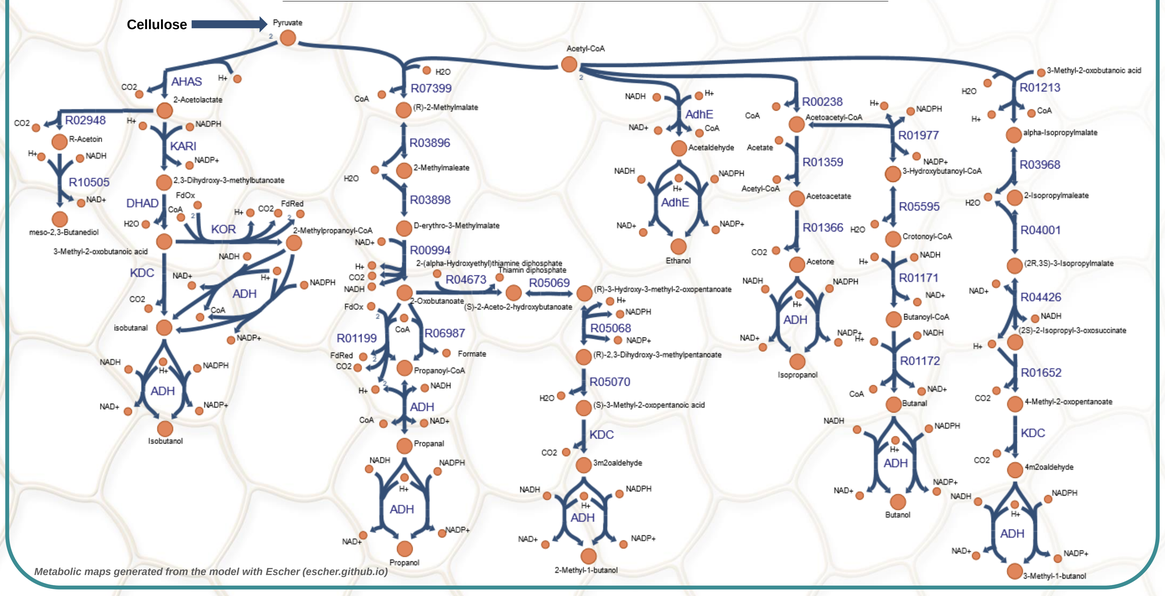
Thompson, R. A., et al. "Exploring complex cellular phenotypes and model-guided strain design with a novel genome-scale metabolic model of Clostridium thermocellum DSM 1313 implementing an adjustable cellulosome." Biotechnology for biofuels 9.1 (2016): 194.

Alcohol Production Modules

We targeted different alcohols with biofuel applications. Most of these alcohols were detected* in the wild type *C. thermocellum.*

** Holwerda, E. K., et al. "The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading." Biotechnology for biofuels 7.1 (2014): 155.*

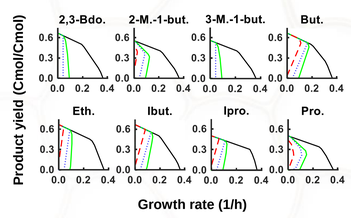
	2,3-butanediol	Isobutanol	Propanol	2-methyl-butanol	Ethanol	Isopropanol	Butanol	3-Methyl-1-Butanol
Detected *in vivo*	+	+	+	+	+	-	-	+
Native enzymes / Total enzymes	1/3	5/7	6/8	8/10	2/2	0/4	3/6	5/7



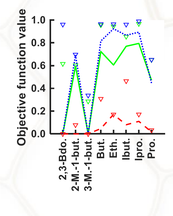
Metabolic maps generated from the model with Escher (escher.github.io)

Results

Feasible metabolic states



Three designs out of many options



wGCP-7-30 wGCP-9-120 sGCP-13-11
Max. for wGCP-7 Max. for wGCP-9 Max. for sGCP-13

Solution encoding: <Objective>-<number of reaction deletions>-<design index>

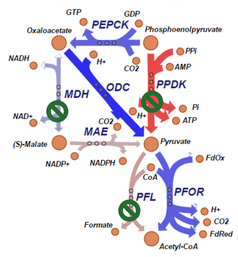
A. Feasible production states of the selected designs. Note that while any point inside the polygon described the axis and plot lines is feasible, the maximum growth rate state coincides with the maximum product yield state. Therefore, the cellular objective and the metabolic engineering objective are coupled.
B. We selected these 3 designs because: (i) They offer wide compatibility (6 out of 8 products) with small trade-off; (ii) the designs can be implemented sequentially, because sGCP-13-11 is a superset of wGCP-9-120 which is a superset of wGCP-7-30. Many other designs were calculated, which optimize different groups of products.

Optimal fluxes (mmol/gCDW/hr) of key metabolic reactions

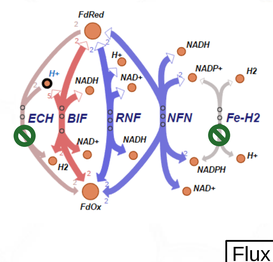
System	Reaction	W.T.	wGCP-7-30					wGCP-9-30					(s)sGCP-13-11				
		Biomass	2m.1but.	But.	Eth.	Ibut.	Ipro.	Pro.	2m.1but.	But.	Eth.	Ibut.	Ipro.	Pro.	2m.1but.	But.	Eth.	Ibut.	Ipro.	Pro.
Redox	ECH	1.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Fe-H2	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	BIF	3.3	1.2	0	0	0	3.6	1.9	3.1	0.5	0.2	0.4	8.6	3.5	3.6	1.3	0.1	0.2	9.1	4.4
	RNF	0.4	-4.7	3.0	1.5	-5.9	-4.8	-1.4	-3.3	4.2	5.2	-6.0	-0.6	-0.6	-1.5	4.9	6.1	-5.8	-0.1	1.6
	NFN	1.2	6.7	6.6	6.8	6.7	4.3	6.4	6.3	6.2	6.3	6.3	3.5	5.9	4.6	5.3	6.2	6.0	3.1	3.6
Pyruvate	PPDK	4.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	PEPCK	6.6	12.2	12.0	12.4	12.1	12.2	12.1	12.4	12.0	12.4	12.2	12.3	12.3	12.8	12.3	12.7	12.5	12.6	12.7
	ODC	3.1	5.8	10.2	8.2	10.6	10.7	4.7	6.3	11.3	12.0	11.6	11.8	5.4	6.7	11.9	12.5	12.1	12.3	6.0
	MDH	-2.3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	MAE	0.8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	PFL	1.4	3.2	1.0	0.6	0.8	8.3	2.6	0	0	0	0	0	0	0	0	0	0	0	0
	PFOR	6.0	0.2	9.1	7.8	0.3	2.5	2.1	3.2	10.7	11.7	0.5	11.5	5.5	3.6	11.5	12.4	0.3	12.1	6.3
Ammonia	GLDH	-2.4	-4.0	-1.9	-4.0	-1.7	-1.5	-4.6	0	0	0	0	0	0	0	0	0	0	0	0
	GOGAT	0	0	0	0	0	0	0	0	0	0	0	0	0	3.21	0.88	0.34	0.69	0.52	3.62
	GOGAT	0	0	0	0	0	0	0	3.59	1.37	0.74	1.08	0.92	4.14	0	0	0	0	0	0
	GS	0.43	0.17	0.25	0.15	0.20	0.17	0.21	3.71	1.60	0.86	1.26	1.07	4.29	3.25	1.04	0.39	0.81	0.61	3.68
	Ammonia uptake	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Urea Uptake	1.49	0.61	0.86	0.51	0.70	0.61	0.72	0.43	0.79	0.43	0.62	0.53	0.53	0.13	0.54	0.19	0.40	0.30	0.22
	Urea amidohydrolase	1.49	0.61	0.86	0.51	0.70	0.61	0.72	0.43	0.79	0.43	0.62	0.53	0.53	0.13	0.54	0.19	0.40	0.30	0.22
Pyrophosphate	PPA	4.61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	ACS	1.36	1.63	1.69	0.73	1.49	0	1.47	1.71	1.66	1.14	1.42	0.00	1.07	0	0	0	0	0	0
	ACK	4.91	1.63	1.69	0.73	1.49	4.01	2.16	1.71	1.66	1.14	1.42	4.28	2.88	0.53	0.96	0	0	5.92	2.92
	PTA	4.91	1.63	1.69	0.73	1.49	4.01	2.16	1.71	1.66	1.14	1.42	4.28	2.88	0.53	0.96	0	0	5.92	2.92
	R01354	2.79	1.22	1.27	0.55	1.12	1.56	1.45	1.29	1.24	0.86	1.06	1.53	1.71	0	0	0	0	0	0
	R00926	2.79	1.22	1.27	0.55	1.12	1.56	1.45	1.29	1.24	0.86	1.06	1.53	1.71	0	0	0	0	0	0
	R00921	2.79	1.22	1.27	0.55	1.12	1.56	1.45	1.29	1.24	0.86	1.06	1.53	1.71	0	0	0	0	0	0
	R01353	2.79	1.22	1.27	0.55	1.12	1.56	1.45	1.29	1.24	0.86	1.06	1.53	1.71	0	0	0	0	0	0
	GC2	1.13	3.14	2.83	2.10	3.31	5.31	2.97	3.13	2.94	4.12	3.49	5.74	3.26	0	0	0	0	0	0
	GC1	1.13	3.14	2.83	2.10	3.31	5.31	2.97	3.13	2.94	4.12	3.49	5.74	3.26	0	0	0	0	0	0
	GLY1	1.13	3.14	2.83	2.10	3.31	5.31	2.97	3.13	2.94	4.12	3.49	5.74	3.26	0	0	0	0	0	0
Other	R04672	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	KOR	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

- The color coding of each reaction corresponds to the additional deletions in each design.
- We calculated the flux distribution at the optimal production state (max growth rate and min product yield), using parsimonious flux balance analysis, which has been demonstrated to be one of the most accurate flux estimation techniques.
- We use the flux distribution form the wild type and sGCP-13-11 in the ethanol production network for the metabolic maps.

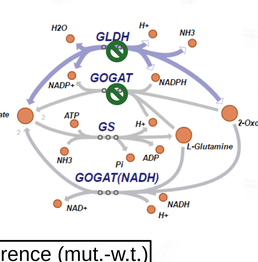
A. Pyruvate Metabolism



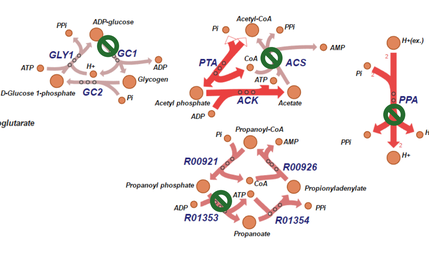
B. Redox Node



C. Ammonia Metabolism



D. Pyrophosphate Metabolism



Flux difference (mut.-w.t.)

A. The avoidance of PPDK is related with Pyrophosphate Metabolism (**E**).
The preference of ODC over MDH+MAE is justified by the requirements and greater flexibility of the Redox Node (**B**).
B. RNF directionality adjusts to the optimum cofactor needs of the target product, and to turnover FdRed from PFOR. To ensure this flexibility NADH must be available. Also, RNF flux in this production strain is 16 fold greater than the wild type.
ECH and Fe-H2 deletions limit the production of hydrogen redirecting electrons towards the target product.
BIF is required to maintain redox homeostasis.
C. These deletions redirect the electrons original intended for amino acid synthesis towards the target product.
D. We hypothesize these deletions constraint growth rate thus redirecting carbon and electrons from biomass towards product.
Several of these reactions are a consequence of too general annotation (e.g. ACS is unlikely to be present since C. therm. does not grow on acetate or consume it; The proponyl-CoA/Propanoate cycle is encoded by the same genes as PTA-ACK- ACS, it is unknown if such reactions occur in C. therm.).
Other (In reference to the table). R04672 is involved in thiamine diphosphate and 2-Acetolactate metabolism. However, this reaction also causes unrealistic behaviors in the model (i.e. thermodynamically infeasible cycles), a common modeling issue which needs to be corrected by further curation.
KOR deletion serves to block isobutanol production in other alcohol production networks. The deletion of KOR in the chassis does not affect the isobutanol production network, because it contains KOR as part of the production module.

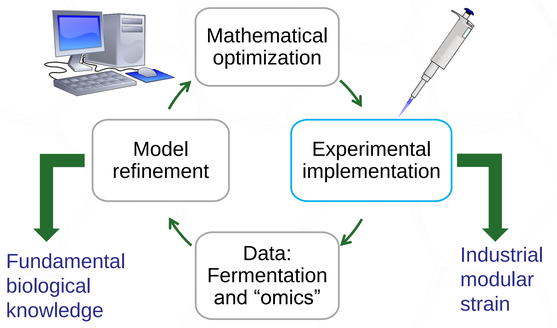
Summary of the designs

- Consistent with experimental evidence from ethanol designs: Constrains in hydrogen and ammonia metabolism, together with RNF overexpression.
- The quantitative analysis of metabolism leads to non-intuitive targets related with pyrophosphate bioenergetic pathways.
- Use of highly flexible Redox Node to allocate electrons into the best cofactor for the current production pathway.

Summary and Future Work

Summary:
- Developed ModCell2, a novel method for modular strain design using multiobjective optimization.
- Demonstrate ModCell2 for two different types of design objectives, weak growth coupled to product formation and strong growth coupled to product formation.
- Use ModCell2 to design *C. thermocellum* modular cells for high yield production of alcohol biofuels.

Future work:
- Model generated hypotheses and possible errors need to be addressed based on literature review and experimentation.



The BioEnergy Science Center (BESC) is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.