

BioLEGO — a web-based application for biorefinery design and evaluation of serial biomass fermentation

Edward Vitkin¹, Alexander Golberg^{*,2} & Zohar Yakhini^{*,1,3}

The composition of feedstock biomass and the selection of fermenting microorganisms are critical factors in biorefinery design. Feedstock biomass composition is constrained by local supply materials, but microorganism selection affords considerable flexibility. Once biomass feedstock is identified, biorefinery designers need to select optimal fermenting organisms. While fermentation by microorganism communities can increase the range of digested biomass compounds and can be more resistant to infections, it has intrinsic problems in the context of species competition, process design and modeling — issues related to insufficient process control. Using a serial fermentation approach, we offset some of these issues to allow maximal process control, while benefiting from organism diversity to maximize feedstock conversion rates. Here, we describe BioLEGO, a freely available web-based application that enables computer-assisted a single and two-step multiorganism fermentation process design. BioLEGO is based on a modular modeling approach, enabling the generation of different fermentation configurations consisting of independent organism modules. BioLEGO supports the evaluation of possible biomass-to-product yields for biomass mixes or general media and recommends media changes to increase the process efficacy.

Keywords: Metabolism Modeling; Serial Fermentation; Biomass; Fermentation Modelling and Optimization; Computer-Assisted Fermentation Modelling; Flux Balance Analysis; Biorefinery Design; Bioethanol from Corn and Seaweeds.

INNOVATION

Although processes in chemical refineries rely on computer-based simulations for decades, the computational simulation and computer-assisted design of fermentation processes are not well developed. The existing models address single-step fermentation configurations only. In this work, we developed a web-based computational service platform, BioLEGO, to simulate the efficiency of biomass fermentation in multi-step processes. Our novel approach to fermentation modeling is serial fermentation particularly relevant for processes that require competing species, or processes with tight control, making modeling of integrated communities less suitable. Briefly, BioLEGO is focused on processes in which each organism is grown separately and then residual media, and the deconstructed organism biomass are transferred for fermentation by the next organism serially. In this work, we provide a novel mathematical modeling approach to designing two-step fermentation processes, a specific case of serial fermentation. BioLEGO enables the assessment of possible biomass-to-product yields for mixtures of biomasses or general media according to various fermentation configurations. It also provides recommendations for media content optimization according to gradient analysis of the computed production rates.

INTRODUCTION

Economically efficient, socially and environmentally sustainable conversion of biomass into valuable products, such as chemicals, food and

fuels, is a major contemporary challenge for science, governments and businesses worldwide¹. Biorefineries integrate all the aspects of biomass processing flow, such as biomass growth, harvesting and fermentation together with distribution of the resulted biofuels and elimination of remaining waste². Clearly, the composition of the feedstock biomass and the ability of microorganisms to efficiently ferment it are two most critical factors influencing the efficiency biorefineries. While the first factor is heavily influenced by available local raw supply materials³, the biorefinery engineers/designers have relative flexibility in designing the fermentation setup. Although processes in chemical refineries heavily rely on computation simulations for decades, the computational simulation and computer-assisted design of fermentation processes are not well developed and the existing models only address single-step fermentation configurations⁴. In this work, we developed an internet-based computational service platform to simulate the efficiency of biomass fermentation in multistep processes.

Once the set of feedstocks possibly available for fermentation is identified, biorefinery designers need to select the organisms for the fermentation process. Natural feedstock biomass is composed of various different molecules, such as monosugars (glucose, galactose, rhamnose, xylose, etc.), amino acids (valine, histidine, lysine, etc.), fatty acids (myristic, oleic, palmitic, etc.), fibers (cellulose, hemicellulose, lignin, ulvan, etc.) and others. Thus, the selection of the fermenting organism is not a trivial issue, since most of the organisms cannot metabolize part of the existing biomass components, leading to significant amount of

^{*}Equal senior author contribution.

¹Department of Computer Science, Technion - Israel Institute of Technology, Haifa, Israel, ²Porter School of Environmental Studies, Tel Aviv University, Tel Aviv, Israel, ³Agilent Laboratories, Tel Aviv, Israel. Correspondence should be addressed to: E.V. (edward.vitkin@gmail.com).

residual media and low fermentation efficiency. Some existing tools^{5–7} are dedicated for simulations of organism feeding trajectories inside bioreactors. Others aim to predict optimal media concentrations to achieve better fermentation yields^{7–9}. For example, wild type *Saccharomyces cerevisiae*, which is a first-choice organism for bioethanol production, poorly utilizes carbohydrates such as xylose, rhamnose and galactose. One approach to overcoming this deficiency and thereby to improving the bioethanol yields is to genetically modify *S. cerevisiae* to improve sugar-uptake mechanisms. Studies in this direction are undertaken for several years but successful implementation remains an open challenge¹⁰. Another approach is to induce or to increase the require functionality in the organism with broader digestion ability. For example, Bond-Watts *et al.*¹¹ proposed different plasmid inserts into *E. coli* to introduce butanol-producing pathways. However, broad digestion ability of some single organisms reduces the total yields of desired products. For example, *E. coli* is less efficient in production of ethanol from glucose than *S. cerevisiae*^{12,13}.

Fermentation by bacterial communities is a natural alternative to genetic modifications of selected organisms. Community members can be selected to naturally digest the broader range of existing biomass compounds and further convert them to desired products. Indeed, fermentation by communities has some serious drawbacks, such as the need for expertise in growing of several organisms, and understanding of inter-organism interactions and competition for resources. Mathematical modeling of the community-based fermentation process is also further complicated, since the natural inter-organism interactions are still poorly understood¹⁴ (relatively to intra-organism metabolism) and since mapping of metabolites between models representing individual organisms is not trivial due to the historical multiplicity of naming conventions^{15,16}. There are several approaches for mathematical modeling of the community behavior. For example, the OptCom¹⁴ methodology proposes a computational framework to describe different intra-species interactions grown together. This framework aims to describe trade-offs between individual vs. community level fitness criteria. Similarly, cFBA¹⁷ is a method that integrates inter-species interactions to achieve maximal growth rate of the entire bacterial community. The SUMEX¹⁸ methodology demonstrates the advantage of maximizing the total molar output exchange minus input exchange of metabolites. As a result, higher SUMEX values correspond to better cellular catabolism ability. Another approach¹⁹ tests the fermentation efficiency of co-culturing species which require different media compounds.

One of the major technical problems in mathematical modeling of bacterial community behavior is mapping between different *in-silico* models of organisms. Metabolic models of organisms are usually created by independent groups of organism experts, each of them with their own labeling conventions. Alternative models developed using automatic reconstructions that enforce the same naming convention over the entire set of models, do not have sufficient quality and act only as initial drafts for further validation by separate expert groups. Thus specifically recognizing that metabolite *A* in *Model1* is the same metabolite as *B* in *Model2*, is a time-consuming and error-prone task.

Another challenge intrinsic to existing metabolic networks is predefined reaction directionality and, specifically, the bi-directional transporter reactions. In some growth conditions, these reactions may insert a target metabolite into the cell, while in others they facilitate its secretion to outside. Thus, connecting two organisms by shared metabolites is problematic, since we cannot make any assumption regarding flux direction of these metabolites. However, if we grow organisms in different bioreactors, we can have a full control on the directionality of the metabolic flux between the organisms. Thus, existing mathematical modeling should be adapted to these realities.

One of the possible approaches to allow for better and predictable control of the inter-organism relations in communities stems from the idea of serial fermentation. In this approach, we trade off the possible benefits of inter-organism interactions, as demonstrated by the above methods, in favor of increased control. Briefly, this means constructing a process

in which each organism is grown separately and then residual media together with resulted grown and decomposed biomass are transferred to the next organism in the chain. This approach has an advantage in the context of broader biomass components utilization, allowing for tighter process control. In addition, serial fermentation removes the demand for community growing expertise (only single-organism growing expertise is required), increases the process flexibility (the modification of one fermenting organism will have low impact on the inter-organism interactions) and simplifies the mathematical process modeling. The simplification of the mathematical modeling is vital in the fermentation design stage since the estimation of expected system efficiency prior to its implementation by simulation will reduce the number of experiments required for process optimization and will provide a new tool for the efficient process design.

In previous work, we have also shown that energy efficient mixing could further improve the net energy balance of the marine biorefinery²⁰. We have also demonstrated the potential of the two-step fermentation approach as a means to increase the yields of ethanol production and carbon utilization from a complex macroalgae *Ulva* biomass²¹. In this paper, we address mathematical modeling aspects of the above-mentioned two-step fermentation process as relevant for processes that require tight control. The implementation of this case is of particular interest, since it will provide practical perspectives on the general serial fermentation process drawbacks and advantages. Specifically, there is a need to evaluate the expected efficiency of proposed single or two-step fermentation processes for the anticipated biomass composition. Mix of biomass composed from different species, such as corn or algae is also of particular interest, since such media can be possibly either be more efficient or cost effective than each single biomass species by itself. Moreover, the composition of the available biomass can be subtly adjusted by changing growing conditions^{22,23}, thus further leading to increased process efficiency.

Here, we present BioLEGO²⁴ — a freely available modeling web service with friendly and totally intuitive interface that enables modeling and evaluation of the expected performance of single and two-step fermentation processes. We propose and implement a flexible modular modeling approach enabling smooth generation of different fermentation configurations consisting of independent encapsulated modules (see ‘Methods’), representing individual organisms. In addition, the BioLEGO web service enables assessment of possible product amounts that can be achieved from specific biomass, mix of biomasses or general media according to various fermentation configurations. Moreover, it provides recommendations for media content improvement according to biomass component gradient analysis (see ‘Methods’) of estimated production rates.

METHODS

Encapsulated microorganism modules

As described above, the mathematical modeling of the community-based fermentation process is complicated due to the lack of understanding of inter-organism interactions and due to nomenclature differences. To offset these issues, we propose a flexible modular modeling approach making the generation of different fermentation configurations akin to combining LEGO²⁵ pieces. In our case, each LEGO piece, called a *module*, contains an encapsulated metabolic model of an individual organism constructed by experts. The underlying metabolic model can be taken from the literature or constructed automatically²⁶. In this section, we address the mathematical aspects of using this approach in formulating biorefinery optimization tasks.

The modular approach – single module

Inspired by the compartmentalized consortium approach²⁷, we address both earlier mentioned problems by encapsulating the existing organism metabolic model in additional (envelope) layer. This leads to sufficiency of mapping between each underlying model and the envelope layer (Fig. 1), thus significantly diminishing the inter-model mapping problem,

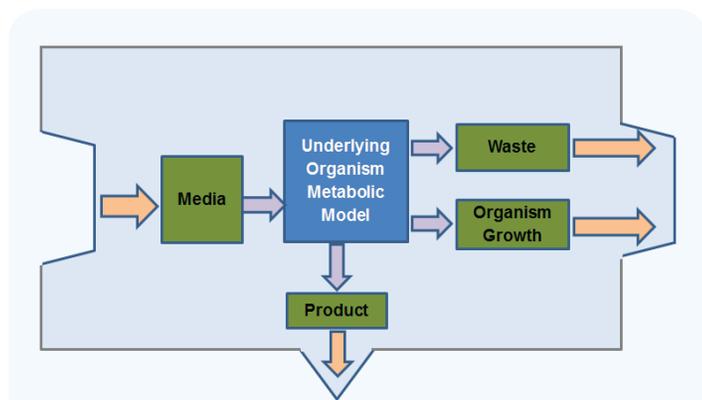


Figure 1 Envelope layer for existing organism metabolic model to create a single module. Each envelope contains five chambers: *Internal* - consisting underlying metabolic model of the organism; *Media* - representing received biomass feedstock media; *Product* - representing the desired target metabolite; *Growth* - consisting organism cellular components; *Waste* - representing all non-digested media residuals and molecules by the organism as growth by-products.

although not completely eliminating it. For the purpose of this mapping, we created new metabolite identifiers based on the proposed convention of KEGG DB²⁸ and continuously updated with the integration of new underlying metabolic models in the BioLEGO website.

For each organism, we define its metabolic model as the *internal* chamber and create an envelope of four external chambers: *media*, *growth*, *waste* and *product*. *Media* refers to the set of media compounds received by organism; *growth* refers to the set of molecules comprising the organism; *waste* refers all non-digested and extracted particles; and *product* refers to the desired product (assuming it can be purified from particles in the *waste* chamber). The inter-chamber metabolic fluxes are defined to be unidirectional, as presented on Fig. 1. Also, note that there is no direct metabolic flux between the *internal* chamber and the outside, which allows full control over all the system inputs and outputs. There is a clear mathematical detachment between the digested metabolite A and extracted metabolite A, although practically it can be the same molecule.

This feature of our modular approach is key to enabling, by combining modules as described below, simple modeling that faithfully represents a controllable process.

The modular approach - combining modules

Once we create the envelope layer around each module (based on the underlying metabolic model), any combination of these envelopes is as simple as combining LEGO²⁵ pieces. All that is needed is to define the organism relationship rules of the desired fermentation process. For example, the serial two-step fermentation process can be enforced by the following rules (Table 1), presented as bold arrows in Fig. 2.

For comparison, the fermentation process in co-culturing of two organisms is enforced by the set of rules presented in Table 2 below. Note that Rule 1.3 transferring 'Growth Org 1' to 'Media Org 2' does not exist in the co-culture system, since cellular components of Organism1 cannot be decomposed in the real time and returned to media. In a co-culture system, we introduce additional possible metabolite flux directions, transferring particles secreted by the second organism as media to the first, as well as providing the entire initial feedstock media to the second organism (Rule 2.7 and Rule 2.6, respectively).

MATHEMATICAL MODELING

Flux balance analysis

Mathematical simulations of biomass utilization and ethanol production yields are performed according to a commonly used Flux

Table 1 Module interaction rules for a two-step fermentation process.

Rule ID	Rule Description (Arrow in Fig. 2)
Rule 1.1	Feedstock Media → Media Org1
Rule 1.2	Waste Org1 → Media Org2
Rule 1.3	Growth Org1 → Media Org2
Rule 1.4	Product Org1 → Total product
Rule 1.5	Waste Org2 → Total waste
Rule 1.6	Growth Org2 → Total waste
Rule 1.7	Product Org2 → Total product

Balance Analysis (FBA) methodology, which is a sub-class of Constraint-Based Modeling (CBM) mathematical modeling frameworks. This method analyzes internal reaction fluxes based solely on simple physical-chemical constraints, such as reaction stoichiometry and metabolic flux constraints, without requiring exact enzyme

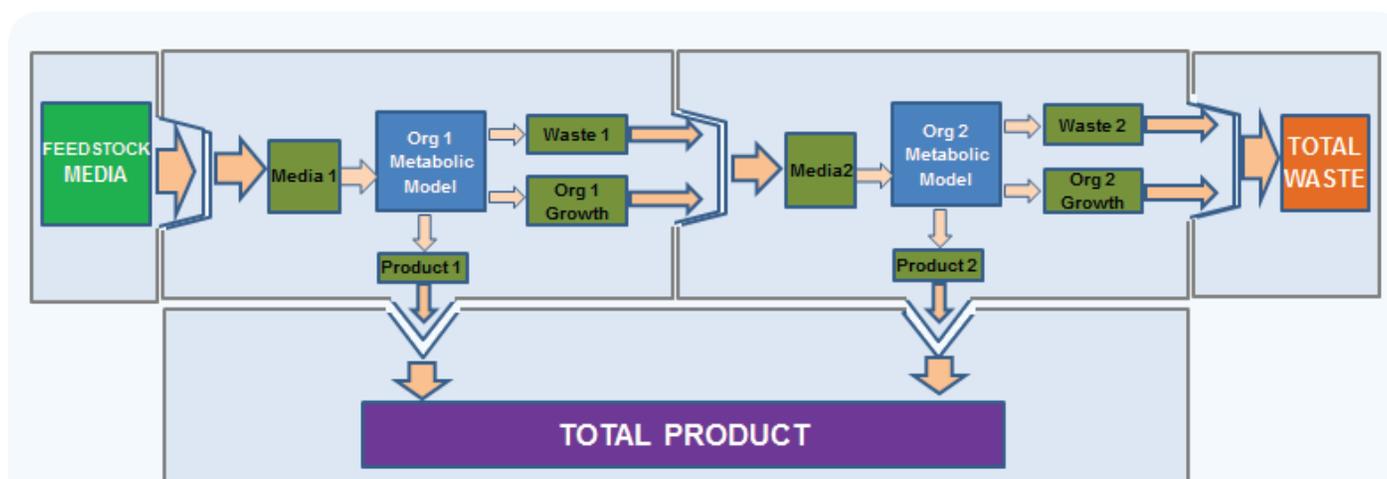


Figure 2 Combination of two single modules to characterize a two-step fermentation process. Three outside modules are added to the process: *Feedstock Media*, *Total Waste* and *Total Product*.

Table 2 Module interaction rules for a single-step fermentation process in co-culturing of two organisms.

Rule ID	Rule Description	Comparison to Table 1
Rule 2.1	Feedstock Media → Media Org1	Same as rule 1.1
Rule 2.2	Waste Org1 → Media Org2	Same as rule 1.2
Rule 2.3	Waste Org1 → Total waste	Novel
Rule 2.4	Growth Org1 → Total waste	Novel
Rule 2.5	Product Org1 → Total product	Same as rule 1.4
Rule 2.6	Feedstock Media → Media Org2	Novel
Rule 2.7	Waste Org2 → Media Org1	Novel
Rule 2.8	Waste Org2 → Total waste	Same as rule 1.5
Rule 2.9	Growth Org2 → Total waste	Same as rule 1.6
Rule 2.10	Product Org2 → Total product	Same as rule 1.7

kinetic data. Specifically, such a methodology enables the prediction of organism growth rates based only on reaction stoichiometry and directionality. FBA-based approaches have a wide range of applications including phenotype analysis, bioengineering and metabolic model reconstructions^{26,29–31}.

There are two constraint types widely used in various CBM-based methods: (i) mass-balance constraints imposed by network stoichiometry and (ii) maximal/minimal feasible reaction flux constraints. Mass-balance constraints dictate that for each metabolite, the sum of fluxes for all reactions producing the metabolite is equal to the sum of fluxes consuming it Eq. (1):

$$S \cdot \vec{r} = 0 \quad (1)$$

where S is a stoichiometric matrix, in which $S_{m,r}$ corresponds to stoichiometric coefficient of metabolite m in the reaction r and \vec{r} is a vector of reaction fluxes (v_r is a metabolic flux through reaction r). Maximal/minimal feasible reaction fluxes (v_r^{LB} and v_r^{UB} in Eq. (2) are usually set to $[-Inf; Inf]$ for bidirectional and $[0; Inf]$ for unidirectional internal reactions:

$$v_r^{LB} \leq v_r \leq v_r^{UB}, \quad \forall r \in \text{reactions} \quad (2)$$

Although the v_r^{LB} and v_r^{UB} are set to infinity for most of reactions, the solution space is not unbound; it is limited by constraining the feedstock media uptake rate. In our specific case, the knowledge of actual media uptake rate is not critical, because we are not interested in reaction rates, but rather in total conversion yield (in %) of biomass into ethanol. Therefore, we assume the uptake rate of 1 gDW \cdot h $^{-1}$ of media (enforced on the media transporter reactions) and calculate the biomass-to-ethanol conversion yield accordingly. Note that such assumption results in rates very similar to the accepted. For example, common glucose uptake rate assumed by *E. coli* is 1.8 gDW \cdot h $^{-1}$ for glucose minimal media³².

FBA is a particular case of the CBM framework, assuming that the modeled organism metabolic network is regulated so as to maximize some cellular function. Organism growth rate is a commonly accepted optimization criterion for unicellular organisms³³. The final FBA formulation is presented in Eq. (3), below:

$$\max_{\vec{v}} \{v_{Growth}\} \text{ s.t. : } \left[\begin{array}{l} S \cdot \vec{r} = 0 \\ v_r^{LB} \leq v_r \leq v_r^{UB} \\ v_{\text{mediatransporters}} = 1 \frac{\text{gDW}}{\text{h}} \text{Media} \end{array} \right], \quad (3)$$

in which v_{Growth} is an artificial growth reaction converting all the organism cellular components into unit of growth. Indeed, there may be many possible sets of fluxes that satisfy all of the CBM constraints and maximize v_{Growth} .

There are several approaches aiming to further constrain the possible flux space represented by Eq. (1,2). For example, TMFA³⁴ provides a methodology to enforce the reaction directionality according to reaction free energy values (ΔG). This method requires prior knowledge of ΔG^0 , of which exact values are known only for a minority of the reactions. The approximation of these values is also possible³⁴. Another approach to limit the flux space is by introducing topological constraints, as presented by FBAwMC methodology³⁵. This method is based on molecular crowding considerations, providing a mathematical framework to limit the amount of molecules which can simultaneously appear in the cell. Thus, this system is limiting the rates of corresponding reactions. Although beneficial, currently these and other additional constraints are not integrated in BioLEGO.

Note that under fluxes that maximize growth rate as described by Eq. (3), each non-growth reaction (e.g., ethanol production) may have a range of possible values. Modeling aims to estimate this range, formally described as Flux Variability Analysis (FVA)³⁶, as presented in Eq. (4):

$$\max / \min \{v_{ethanol}\} \text{ s.t. : } \left[\begin{array}{l} S \cdot \vec{r} = 0 \\ v_r^{LB} \leq v_r \leq v_r^{UB} \\ v_{\text{mediatransporters}} = 1 \frac{\text{gDW}}{\text{h}} \text{Media} \\ v_{Growth} \text{ is a maximal solution of Eq. (3)} \end{array} \right]. \quad (4)$$

In BioLEGO, we perform evaluation steps under this general framework and combine them. All the linear optimizations are performed using GNU Linear Programming Kit³⁷ for solving linear programming problems.

Modeling a single encapsulated module

Several adjustments to the existing CBM constraints Eq. (1,2) are required to create the encapsulation of the existing metabolic models as described above (Fig. 1). First, the encapsulation of the underlying organism stoichiometric model (S -matrix from Eq. (1)) is created (Fig. 3), resulting S_{Env} matrix.

For each exterior non-product metabolite existing in the underlying metabolic model of organism or in the fermented media, we add rows both to *media* and to *waste* chambers. Another row(s) is created for the desired product as *product* chamber. Finally, additional rows are added for each cellular content component to construct *growth* chamber. Once all novel metabolites are introduced into the matrix, we add reaction columns corresponding to the permitted inter-chamber flux directionality (Fig. 1). For example (Fig. 3, *Internal* → *Growth* column), we add reactions transferring cellular components to the metabolites of the *Growth* chamber, which represents the decomposition process of organism cellular components possibly existing between the bioreactors. In addition, temporary columns are added to simulate metabolite exchange with the environment. These columns are omitted in the construction of the two-step fermentation process. The upper and lower boundaries for all the newly added inter-chamber reactions in the system are set to $[0; Inf]$, thus mathematically enforcing existing inter-chamber metabolic flux.

To estimate the fermentation characteristics of a single module, i.e., the fermentation potential of a single organism, the bi-level FVA³⁶ framework is formulated as described in Eq. (5) below:

$$\begin{array}{l}
 \max / \min \{ \text{Total Product} \equiv v_{\text{ethanol}} \} \\
 \left. \begin{array}{l}
 \text{s.t.:} \\
 v_{\text{Growth}} \text{ is a maximal solution of} \\
 \max_{\substack{r \\ v}} \{ v_{\text{Growth}} \} \text{ s.t.:} \\
 [\text{CBM Constraints} = \{ \dots \} - \text{Same as (*)}]
 \end{array} \right\}
 \end{array}
 \left[\begin{array}{l}
 \text{CBM Constraints} = \left\{ \begin{array}{l}
 S_{\text{Env}} \cdot \substack{r \\ v} = 0 \\
 v_r^{\text{LB}} \leq v_r \leq v_r^{\text{UB}} : \forall r \in \text{Originally existing model reactions} \\
 0 \leq v_r \leq \text{Inf} : \forall r \in \text{Novel inter - chamber reactions} \\
 v_{\text{IN Feedstock Media}} = 1 \frac{g^{\text{DW}}}{h} \text{Media}
 \end{array} \right\} (*)
 \end{array} \right. \quad (5)$$

Basically, this equals to substituting S_{Env} into the general variable S and to the addition of inter-chamber reactions boundaries in Eq. (4).

Modeling two-step fermentation process

Simulating the two-step fermentation process adds another complexity level. We have developed a mathematical modeling formulation and framework for this task, as follows. First, we combine both used organism modules into the single stoichiometric system, described by S_{System} matrix (Fig. 4) according to the rules describing the two-step fermentation process (Fig. 2 and Table 1).

The matrix S_{System} is created by appending S_{Env} modules for both organisms (while removing columns describing interactions with the environment) and addition of columns for reaction-transferring metabolites according to predefined rules. Columns describing metabolic exchanges with the environment are added outside of organism modules as presented on Fig. 4. Again, we set the reaction boundaries for all newly added reactions are set to $[0; \text{Inf}]$, similar to inter-chamber case.

The resulting FVA framework is updated from the bi-level to the tri-level formulation, as described by Eq. (6):

$$\begin{array}{l}
 \max / \min \{ \text{Total Product} \equiv v_{\text{Org1-ethanol}} + v_{\text{Org2-ethanol}} \} \\
 \left. \begin{array}{l}
 \text{s.t.:} \\
 v_{\text{Org1-Growth}} \text{ is a maximal solution of : } \max_{\substack{r \\ v}} \{ v_{\text{Org1-Growth}} \} \text{ s.t.:} \\
 \text{CBM constraints} = \{ \dots \} - \text{Same as (**)} \\
 v_{\text{Org1-Growth}} \text{ is a maximal solution of : } \max_{\substack{r \\ v}} \{ v_{\text{Org1-Growth}} \} \text{ s.t.:} \\
 \left[\begin{array}{l}
 \text{CBM constraints : } \{ \dots \} - \text{Same as (**)} \\
 v_{\text{Org1-Growth}} \text{ is a maximal solution of : } \max_{\substack{r \\ v}} \{ v_{\text{Org1-Growth}} \} \text{ s.t.:} \\
 [\text{CBM constraints : } \{ \dots \} - \text{Same as (**)}]
 \end{array} \right]
 \end{array} \right\}
 \end{array}
 \left[\begin{array}{l}
 \text{CBM Constraints} = \left\{ \begin{array}{l}
 S_{\text{System}} \cdot \substack{r \\ v} = 0 \\
 v_r^{\text{LB}} \leq v_r \leq v_r^{\text{UB}} : \forall r \in \text{Originally existing model reactions} \\
 0 \leq v_r \leq \text{Inf} : \forall r \in \text{Inter - chamber \& Inter - module reactions} \\
 v_{\text{IN Feedstock Media}} = 1 \frac{g^{\text{DW}}}{h} \text{Media}
 \end{array} \right\} (**)
 \end{array} \right. \quad (6)$$

Namely, initially we estimate the maximal growth rate of the first organism in the fermentation process according to the given feedstock media and the stoichiometric constraints (S_{System}). Next, we not only estimate the maximal growth rate of the second organism under the same feedstock media and S_{System} , but also limiting the growth rate of the first to the calculated maximum. And finally, we estimate maximal and minimal total ethanol production rates under original feedstock media and stoichiometry constraints together with limiting growth rates of first and second organisms to the calculated values.

Assessment of media changes and the calculation of media gradients

The optimizations of the input media composition are performed using the gradient of the product relative to the media. The component i of the gradient direction vector for function f at some arbitrary point x_0 is defined according to Eq. (7):

$$\nabla f_i(x_0) = \lim_{\delta \rightarrow 0} \left\{ \frac{f(x_0 + \delta \cdot e_i) - f(x_0)}{\delta} \right\}, \quad (7)$$

Chamber Name	Metabolite Type	Original RXNs	Inter-chamber groups of reactions					Interactions with environment				
			Media → Internal	Media → Waste	Internal → Waste	Internal → Product	Internal → Growth	OUTSIDE → Media	Waste → OUTSIDE	Product → OUTSIDE	Growth → OUTSIDE	
$S_{Env} =$ Internal [Original mets]	Internal mets	Original S matrix										
	External non-product mets		I		-I							
	Product mets						-I					
	Growth mets							-I				
Media [Mets received as feedstock]	Existing in external non-product mets		-I						I			
	NOT Existing in external non-product mets			-I								
Waste	Waste mets			I	I					-I		
Product	Product mets					I					-I	
Growth	Growth mets						I					-I

Figure 3 Construction of S_{Env} matrix. **I** stand for the unitary matrix of appropriate dimension (varies from organism to organism). All empty matrix cells are equal to 0, which is not shown for visual compactness.

		Intra-module reactions		Inter-module groups of reactions (Rules 1.1 – 1.7)							Interactions with environment			
		Org 1	Org 2	1.1	1.2	1.3	1.4	1.5	1.6	1.7	OUTSIDE → Media	Waste → OUTSIDE	Product → OUTSIDE	
$S_{System} =$	Feedstock Media			-I								I		
	Module of Org 1	Media	S_{Env}		I									
		Waste			-I									
		Growth				-I								
Module of Org 2	Product					-I								
	Media	S_{Env}		I	I									
	Waste						-I							
Growth							-I							
	Product								-I					
	Total Waste							I	I			-I		
	Total Product						I			I			-I	

Figure 4 Construction of S_{System} matrix. **I** stands for the unitary matrix of appropriate dimensions (varies from organism to organism). All empty matrix cells are equal to 0, which is not shown for visual compactness. For the same reason we omit *internal* chambers and all intra-module reactions described in **Fig. 3**, because they do not contribute to inter-module relationships.

where e_i denotes the i -th unit vector: its i -th entry is one, all others are zero. In our case, x_0 is the originally received media content and f is the estimated ethanol production rate operating on media input by applying our model **Eq. (6)**.

To calculate the media gradient we first calculate the $f(x_0)$ according to **Eq. (6)**. Then, we iteratively increase the amount of each media

particle (each particle i corresponds to the i -th component of the vector x_0) by small δ (currently pre-set to 0.01 mg) and estimate the gradient vector $\nabla f(x_0)$.

Once gradient direction is estimated, the media composition is updated by an α -step (varying in range (0;1] while enforcing the media vector to non-negative values) in this direction, which is the direction of maximal product increase, while keeping the total media weight constant **Eq. (8)**.

$$\text{Update: } \bar{x}_1 \leftarrow \left[\bar{x}_0 + \alpha \nabla f(\bar{x}_0) \right] \quad (8a)$$

$$\text{Normalization: } \bar{x}_1 \rightarrow \bar{x}_1 \frac{x_0}{x_1} \quad (8b)$$

The process is iteratively repeated for the new point until convergence to the local maximum is achieved or maximal number of iterations is reached.

This method is targeted to identify slight media adjustments that can be a result either of changes in fermentation fertilizers or of changes in feedstock biomass growing condition. To evaluate the expected efficiency associated with

more significant media changes obtained by mix of different feedstocks, BioLEGO provides the dedicated option in the user interface.

RESULTS

The BioLEGO web-based application

The BioLEGO web service is dedicated to create a user-friendly web interface (**Fig. 5**) to estimate the expected efficiency of different fermentation setups for the selected complex media, such as decomposed corn or algal biomass. It does not require any special programming skills or other similar expertise from the user, neither does it require any time and resource investments in the installation of MATLAB³⁸ or dedicated metabolism simulating tools. Moreover, it provides several novel unique simulation capabilities, such as the simulation of two-step fermentation process and optimization of input media.

In order to evaluate the fermentation configuration for the media of interest, the user needs to perform several simple steps. First, he needs to provide the relevant media as input, which can be done by the addition of each individual molecule and its relative quantity (**Fig. 5e**) or by selecting one or mix of two predefined media and then adjusting its content (**Fig. 5a**). Next, user should choose the desired acting organisms and their fermentation configuration (**Fig. 5c**). Several configurations can

be tested simultaneously. Finally, the user should select the requested output product (**Fig. 5b**). Although the current numbers of predefined media, fermenting organisms and target products are still not significant, BioLEGO website is constantly updated with new options. More details on BioLEGO options can be found on the website: <http://bioinfo.cs.technion.ac.il/people/zohar/BioLEGO>.

BioLEGO

(a) **Input Media Composition**
 First default media: U. lactuca (60%)
 Second default media: Kappaphycus alvarezii w/o fibers (40%)

(b) **Product**
 Ethanol

(c) **Organisms**
 Organism 1: E. coli
 Organism 2: S. cerevisiae WT

(d) **Options:**
 Aerobic
 Actual Media Uptake
 Maximal Product Media Gradient
 Minimal Product Media Gradient

(e) **Configuration**
 Single-Step Fermentation: Org1, Org2, Co-culture
 Two-Step Fermentation: Org1—Org2, Org2—Org1

Media Components and Quantities:

Category	Component	Quantity [g/L]
Carbohydrates	D-galactose	59.6
	3,6-anhydrogalactose	44.0
	D-glucose	33.8
	D-xylose	2.0
Fatty Acids	Myristic acid	1.36100
	Myristoleic acid	7.13300
	Pentadecanoic acid	1.42080
	Palmitoleic acid	5.17200
Amino Acids	L-aspartate	2.3
	L-glutamate	4.5
	L-asparagine	4.2
	L-serine	0.4

Nucleotides

Minerals

Unclassified metabolites	L-Arabinose	0.4
	Vitamin A1	3.47900
	Other particles	201.7

Version 1.0 from March 11, 2015
 © Edward Veljkovic

Figure 5 BioLEGO website: <http://bioinfo.cs.technion.ac.il/people/zohar/BioLEGO> (a) Selection of default input feedstock media, (b) selection of target product, (c) choice of fermentation configuration, (d) additional input and output options and (e) list of selected media components and their quantities.

Input Media Composition
 First default media: Corn Cobs (100%)
 Second default media: --- (0%)

Product
 Ethanol

Options:
 Aerobic
 Actual Media Uptake
 Maximal Product Media Gradient
 Minimal Product Media Gradient

Organisms
 Organism 1: S. cerevisiae WT
 Organism 2: S. cerevisiae RN1016 (+ Xylose isomerase)

Configuration
 Single-Step Fermentation: Org1, Org2, Co-culture
 Two-Step Fermentation: Org1—Org2, Org2—Org1

Output Summary:

Configuration	<i>S. cerevisiae</i> WT growth [h ⁻¹]	<i>S. cerevisiae</i> RN1016 (+ Xylose isomerase) growth [h ⁻¹]	Min Ethanol		Max Ethanol		Details
			Production [g/Kg]	Carbon Utilization [%]	Production [g/Kg]	Carbon Utilization [%]	
<i>S. cerevisiae</i> WT	0.0498	---	132.4835	20.9%	164.8431	26.0%	>>>
<i>S. cerevisiae</i> RN1016 (+ Xylose isomerase)	---	0.0642	268.6756	42.4%	300.8463	47.5%	>>>

Figure 6 BioLEGO run for comparison of corn cobs single-step fermentation efficiency by WT and RN1016 yeasts. (a) Input configuration and (b) output summary.

System usage examples

In this section, we describe two illustrative scenarios for BioLEGO usage.

Biomass optimization and single-step fermentation of corn cobs by WT and RN1016 yeast

First, suppose we want to compare single-step fermentation efficiency of corn cobs between two different yeast species: wild type (WT) and recombinant (RN1016), which has a *Piromyces xylose-isomerase*³⁹ in its genome that facilitates the fermentation of both glucose and xylose to ethanol. Moreover, we want to test the potential of slight corn cob biomass changes to enhance this process. As expected, the input to the system will be the choice of a single default medium = “Corn Cobs” and both yeast species. In addition, one needs to select calculation of gradients for maximal and minimal ethanol production and check all relevant fermentation configurations (Fig. 6a).

The simulation for both species should take less than 10 minutes (average of three runs is 5.54) on the current server (six cores of 2.53 GHz, 8 GB total memory). Results (Fig. 6b) show that WT yeast should produce 132.5–164.8 g ethanol per 1 kg of corn cobs (g/kg). These estimations are well correlated with the experimental results demonstrated by Eylen *et al.*⁴⁰, where yield of 158 g/kg was demonstrated. On the other hand, for the RN1016 yeast we predict a production rate of 268.7–300.8 g/kg, which is significantly higher than the reported 197 g/kg⁴⁰. This disagreement is expected because, as we mentioned earlier, the assumption of highly efficient xylose digestion efficiency by *S. cerevisiae* is still problematic¹⁰. To fix this issue, the underlying metabolic model of RN1016 yeast can be refined by limiting the rate of xylose-isomerase reaction to fit the experimental results reported above. Indeed, such an update would require sufficient understanding both of FBA modeling methodology as well as of yeast metabolic behavior.

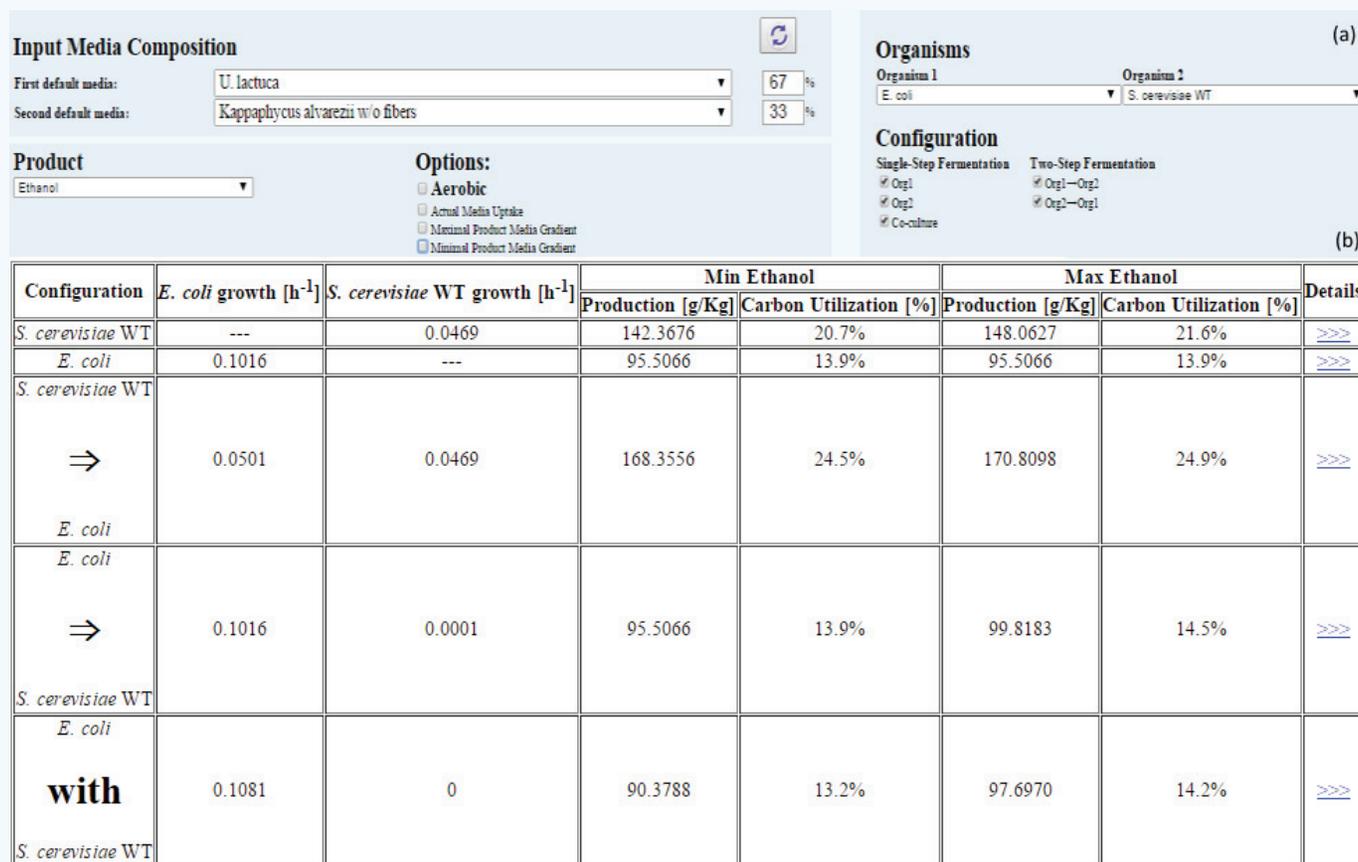


Figure 7 BioLEGO run for evaluation of efficiency of the two-step fermentation process for 2:1 mix of *U. lactuca* together with *K. alvarezii* biomasses by *E. coli* and WT *S. cerevisiae*. (a) Input configuration and (b) output summary.

The analysis of possibly improving corn cob biomass composition yields some expected results and some that we less expected absent the detailed calculations (Supplementary Fig. 1). As expected, our simulations suggest decreasing the relative weight of all non-digested particles, while increasing glucose or glucose and xylose relative weights for WT and for RN1016 yeasts, respectively. Interestingly, system suggestions for maximizing minimal versus maximal ethanol productions differ for several amino acids. For example, for maximizing minimal ethanol production, our system proposes to increase the proportion of valine, lysine and isoleucine, while it is unnecessary for maximizing maximal ethanol production.

Two-step fermentation of 2:1 mix of *U. lactuca* and *K. alvarezii* biomasses by *E. coli* and WT *S. cerevisiae*

The second case scenario we describe here is the evaluation of efficiency of a two-step fermentation process for 2:1 mix of *U. lactuca* and *K. alvarezii* algal biomasses by *E. coli* and WT *S. cerevisiae*. We would like to estimate both configuration directions, i.e., *E. coli* followed by yeast and vice versa. In addition, we want to compare the efficiency of the two-step fermentation process to the single-step fermentation by each organism separately and by the organism co-culture. The last configuration is created according to the rules presented in Table 2. To achieve all these targets, the input to the system will be the choice of two default media: “*U. lactuca*” and “*K. alvarezii*” and both desired organisms. In addition, one needs to provide the ratio of the selected biomasses and check all relevant fermentation configurations (Fig. 7a).

The simulation for all five configurations should take less than one minute (average of three runs is 31 seconds) on the current server (six cores of 2.53 GHz, 8 GB total memory). Simulation results (Fig. 7b) predict that two-step fermentation starting with WT yeast should produce 168.4–170.8 g ethanol per 1 kg of media mix, which has clear advantage over all other scenarios, demonstrating ethanol production yield of 142.4–148.1 g of ethanol for WT *S. cerevisiae* alone or around 90–100 g of ethanol for *E. coli* consisting scenarios. Note that the last value includes evaluation of the co-culturing setup estimation, predicting 90.4–97.7 g of ethanol per 1 kg of media mix.

This use case therefore demonstrates the usefulness of our modeling approach in designing a fermentation process.

DISCUSSION

Computational and modeling issues and limitations

Several aspects should be considered when using BioLEGO and applying its results. One issue is the mapping between different naming conventions of incorporated underlying metabolic models. Although this problem was significantly diminished by the introduction of our envelope layer, there still exists a mapping of metabolites between each underlying metabolic model to the BioLEGO envelope and such mapping can be imprecise due to nomenclature differences.

Another inherent issue is the quality of underlying metabolic models for fermenting organisms. Indeed, these models were created and validated by experts, but there is a non-negligible gap between the actual

metabolic behaviors in the organisms and the modeled ones. Moreover, many organisms may require additional media molecules (i.e., water) or growth factors (i.e., ergosterol). Most of these molecules are already incorporated into the underlying organism models, but may not exist in the estimated media provided by the user. We overcome this issue by the inclusion of the required molecules into the media.

Additional and orthogonal inconsistencies may stem from the mathematical nature of simulations. One problem is the variability in reaction stoichiometric coefficients existing in underlying metabolic models. This variability can reach eight orders of magnitude, which may lead to numerical problems in the linear solver. Another problem is the assumption of the cellular optimization target. We calculate the production efficiency under the assumption of maximal organism growth rate, while in practice this may not always be true. Providing an option to estimate the production efficiency values while permitting suboptimal organism growth rate is among the future directions for BioLEGO development.

Finally, the chemical composition of a feedstock compiled from various literature sources is collective, and thus should be cautiously used. The experiments provided by these sources may differ in biomass analysis methods, in biomass growing conditions and in selected specific biomass species. Hence, to achieve more accurate results, one needs to estimate and provide the exact media composition of feedstock biomass.

Future directions

Indeed, the BioLEGO web-based application is a living on-going project. Several additional features are expected to be integrated to the project in later stages.

One such feature is the evaluation of media fermentation performance by a proposed system configuration together with genetic modifications of the participant organisms. Indeed, automatic modeling of organism metabolic behavior change due to induced functionality is problematic, but such modeling due to gene knockouts is already possible. Thus, we intend to integrate into the system (i) evaluation of system performance resulting from the deletion of specific user-defined gene or set of genes and (ii) identification of optimal set of genes to be knocked out.

Another expected system improvement is the generalization of the web service to three organism fermentation processes. This introduces another level of model complexity, since the number of configuration setups for three organisms grows significantly. Not only three-step lineal fermentation processes can be taken into consideration, but also two-step brunching ones, where after the first organism, the unutilized media particles are split, and some are forwarded to one organism and some to another.

The estimation of potential of designing a multiproduct system can also be developed. This is a very promising direction, since if we physically separate the fermenting organisms into different bioreactors, we can adapt the process to produce different biomolecules. Also, following the same reasoning, different media fertilizers for each organism can be added separately, thus enhancing its bioproduction efficiency. As a result, more complex organism feeding setups can be considered. In addition, the already existing estimation of potential system performance for different media mixes will be enhanced to optimize the ratio of the considered media.

Finally, the overall system user interface is expected to develop to enhance the system usability and utility, for example by uploading organism models and feedstock media compositions in SBML or XLS files, and to connect to the existing biomass composition databases.

SUMMARY

The BioLEGO²³ website is freely available for use by the community. As an efficient web-based application and with future improvements, we hope that BioLEGO will prove useful in designing more efficient fermentation processes.

ACKNOWLEDGEMENTS

We would like to thank the anonymous reviewers for valuable and helpful comments. We would like to thank the Computer Science Department of Technion — Israel Institute of Technology for providing the hardware for the BioLEGO website.

REFERENCES

- Karp, A. & Richter, G.M. Meeting the challenge of food and energy security. *J. Exp. Bot.* **62**, 3263–3271 (2011).
- Star-COLIBRI. European Biorefinery Joint Strategic Research Roadmap for 2020 (http://beaconwales.org/uploads/resources/Vision_2020_European_Biorefinery_Joint_Strategic_Research_Roadmap.pdf, 2011).
- Demirbas, M.F. Biorefineries for biofuel upgrading: A critical review. *Appl. Energ.* **86**, S151–S161 (2009).
- Hanly, T.J. & Henson, M.A. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.* **108**, 376–385 (2011).
- Mendes, R., Rocha, I., Ferreira, E.C. & Rocha, M.A. Comparison of algorithms for the optimization of fermentation processes. *Proc. IEEE Int. Conf. Evol. Comput.* 2018–2025 (2006).
- Rocha, M., Mendes, R., Rocha, O., Rocha, I. & Ferreira, E.C. Optimization of fed-batch fermentation processes with bio-inspired algorithms. *Expert Syst. Appl.* **41**, 2186–2195 (2014).
- Rocha, O., Maia, P., Rocha, I. & Rocha, M. OptFerm — a computational platform for the optimization of fermentation processes [Internet]. *Proc. Eur. Simul. Model. Conf.* 5–9 (2014).
- Beg, Q.K., Sahai, V. & Gupta, R. Statistical media optimization and alkaline protease production from *Bacillus mojavensis* in a bioreactor. *Process Biochem.* **39**, 203–209 (2003).
- Zhao, X., Kong, X., Hua, Y., Feng, B. & Zhao, Z. Medium optimization for lipid production through co-fermentation of glucose and xylose by the oleaginous yeast *Lipomyces starkeyi*. *Eur. J. Lipid Sci. Technol.* **110**, 405–412 (2008).
- Van Maris, A.J. et al. Alcoholic fermentation of carbon sources in biomass hydrolysates by *Saccharomyces cerevisiae*: Current status. *Antonie Van Leeuwenhoek* **90**, 391–418 (2006).
- Bond-Watts, B.B., Bellerose, R.J. & Chang, M.C.Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* **7**, 222–227 (2011).
- Kim, Y., Ingram, L.O. & Shanmugam, K.T. Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. *Appl. Environ. Microbiol.* **73**, 1766–1771 (2007).
- Talebna, F., Niklasson, C. & Taherzadeh, M.J. Ethanol production from glucose and dilute-acid hydrolysates by encapsulated *S. cerevisiae*. *Biotechnol. Bioeng.* **90**, 345–353 (2005).
- Zomorodi, A.R. & Maranas, C.D. OptCom: A multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput. Biol.* **8**, e1002363 (2012).
- Durink, S., Spellman, P.T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).
- Van Iersel, M.P. et al. The BridgeDb framework: Standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics* **11**, 5 (2010).
- Khandelwal, R.A., Olivier, B.G., Röling, W.F.M., Teusink, B. & Bruggeman, F.J. Community flux balance analysis for microbial consortia at balanced growth. *PLoS One* **8**, e64567 (2013).
- Zarecki, R. et al. Maximal sum of metabolic exchange fluxes outperforms biomass yield as a predictor of growth rate of microorganisms. *PLoS One* **9**, e98372 (2014).
- Hanly, T.J. & Henson, M.A. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol. Bioeng.* **108**, 376–385 (2011).
- Goldberg, A. & Liberzon, A. Modeling of smart mixing regimes to improve marine biorefinery productivity and energy efficiency. *Algal Res.* **11**, 28–32 (2015).
- Golberg, A., Vitkin, E., Linshiz, G., Khan, S.A., Hillson, N.J., Yakhini, Z., & Yarmush, M.L. Proposed design of distributed macroalgal biorefineries: Thermodynamics, bioconversion technology, and sustainability implications for developing economies. *Biofuels Bioprod. Biorefin.* **8**, 67–82 (2014).
- Suresh Kumar, K., Ganesan, K. & Subba Rao, P.V. Seasonal variation in nutritional composition of *Kappaphycus alvarezii* (Doty) Doty — an edible seaweed. *J. Food Sci. Technol.* **52**, 2751–2760 (2014).
- Marinho-Soriano, E., Fonseca, P.C., Carneiro, M.A.A. & Moreira, W.S.C. Seasonal variation in the chemical composition of two tropical seaweeds. *Bioresour. Technol.* **97**, 2402–2406 (2006).
- BioLEGO Website. [Internet] <http://bioinfo.cs.technion.ac.il/people/zohar/BioLEGO>
- LEGO Website. Home [Internet] <http://www.lego.com/en-us/>

26. Vitkin, E. & Shlomi, T. MIRAGE: A functional genomics-based approach for metabolic network model reconstruction and its application to cyanobacteria networks. *Genome Biol.* **13**, R111 (2012).
27. Taffs, R. *et al.* *In silico* approaches to study mass and energy flows in microbial consortia: A syntrophic case study. *BMC Syst. Biol.* **3**, 114 (2009).
28. Ogata, H. *et al.* KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
29. Price, N.D., Papin, J.A., Schilling, C.H. & Palsson, B.O. Genome-scale microbial *in silico* models: The constraints-based approach. *Trends Biotechnol.* **21**, 162–169 (2003).
30. Stelling, J. *et al.* Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**, 190–193 (2002).
31. Price, N.D., Reed, J.L. & Palsson, B.O. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* **2**, 886–897 (2004).
32. Orth, J.D. *et al.* A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism — 2011. *Mol. Syst. Biol.* **7**, 535 (2011).
33. Raman, K. & Chandra, N. Flux balance analysis of biological systems: Applications and challenges. *Brief. Bioinformatics* **10**, 435–449 (2009).
34. Henry, C.S., Broadbelt, L.J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805 (2007).
35. Beg, Q.K. *et al.* Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12663–12668 (2007).
36. Mahadevan, R. & Schilling, C.H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
37. GLPK - GNU Project - Free Software Foundation (FSF). [Internet] <http://www.gnu.org/software/glpk/>
38. MATLAB - The Language of Technical Computing. [Internet] <http://www.mathworks.com/products/matlab/>
39. Harhangi, H.R. *et al.* Xylose metabolism in the anaerobic fungus *Piromyces* sp. strain E2 follows the bacterial pathway. *Arch. Microbiol.* **180**, 134–141 (2003).
40. Van Eylen, D., van Dongen, F., Kabel, M. & de Bont, J. Corn fiber, cobs and stover: Enzyme-aided saccharification and co-fermentation after dilute acid pretreatment. *Bioresour. Technol.* **102**, 5995–6004 (2011).

SUPPLEMENTARY INFORMATION

Gradient maximizing MINIMAL target production

Target production	L-Rhamnose	l-isoleucic acid	L-Lysine	L-Leucine	Other particles	Lipin	oleate	D-Glucuronate	D-xylose	L-Valine	L-histidine	L-phenylalanine	Myristic acid	D-Mannose	D-galactose	L-arabinose	L-tyrosine	L-arginine	D-Glucose	Acetate	palmitate	stearate	L-Threonine	L-Isoleucine	L-tryptophan	3-Methoxy-4-hydroxy-trans-cinnamate	L-Methionine	Total Weight [mg]
132.478	4.00	17.10	0.50	2.40	39.70	170000	9.20	18.00	280.00	1.10	0.50	1.10	0.06	4.00	8.00	24.00	0.90	1.20	355.00	43.00	3.00	0.90	0.80	0.80	0.20	11.00	0.50	999.961
134.639	3.97	16.92	0.51	2.38	39.61	169600	9.10	17.84	278.18	1.11	0.50	1.09	0.06	4.04	7.94	23.84	0.89	1.22	361.66	42.40	2.97	0.89	0.80	0.81	0.20	10.91	0.50	999.961
136.790	3.94	16.74	0.52	2.37	39.51	169200	9.01	17.72	276.34	1.13	0.49	1.08	0.06	4.08	7.88	23.68	0.88	1.24	365.33	41.80	2.94	0.88	0.80	0.82	0.20	10.83	0.51	999.961
138.937	3.91	16.57	0.52	2.35	39.42	168800	8.91	17.57	274.50	1.14	0.49	1.07	0.06	4.12	7.82	23.52	0.88	1.26	368.99	41.21	2.91	0.87	0.80	0.82	0.21	10.74	0.51	999.961
141.079	3.89	16.39	0.53	2.33	39.32	168399	8.81	17.43	272.64	1.15	0.48	1.06	0.06	4.16	7.76	23.36	0.88	1.27	372.66	40.63	2.88	0.86	0.80	0.83	0.21	10.65	0.51	999.961
143.221	3.86	16.21	0.54	2.31	39.23	167999	8.72	17.28	270.78	1.17	0.47	1.05	0.06	4.20	7.70	23.20	0.88	1.29	376.32	40.05	2.86	0.85	0.80	0.84	0.21	10.56	0.51	999.961
145.359	3.83	16.03	0.55	2.30	39.14	167599	8.62	17.14	268.91	1.18	0.47	1.05	0.06	4.24	7.64	23.03	0.88	1.31	379.99	39.47	2.83	0.84	0.80	0.85	0.21	10.47	0.52	999.961
147.492	3.80	15.85	0.56	2.28	39.04	167199	8.52	16.99	267.04	1.19	0.47	1.04	0.06	4.28	7.58	22.87	0.88	1.33	383.66	38.89	2.80	0.83	0.80	0.86	0.21	10.39	0.52	999.961
149.616	3.77	15.67	0.57	2.26	38.95	166799	8.43	16.85	265.18	1.21	0.46	1.03	0.06	4.32	7.52	22.71	0.88	1.34	387.33	38.31	2.77	0.82	0.80	0.87	0.22	10.30	0.52	999.961
151.739	3.74	15.49	0.58	2.24	38.85	166399	8.33	16.70	263.32	1.22	0.46	1.02	0.06	4.36	7.45	22.54	0.89	1.36	391.00	37.72	2.74	0.81	0.80	0.88	0.22	10.21	0.53	999.961
153.848	3.71	15.31	0.59	2.22	38.76	165999	8.23	16.56	261.45	1.24	0.45	1.01	0.06	4.40	7.39	22.38	0.89	1.38	394.67	37.13	2.71	0.80	0.80	0.89	0.22	10.12	0.53	999.961

Gradient maximizing MAXIMAL target production

Target production	L-Rhamnose	l-isoleucic acid	L-Lysine	L-Leucine	Other particles	Lipin	oleate	D-Glucuronate	D-xylose	L-Valine	L-histidine	L-phenylalanine	Myristic acid	D-Mannose	D-galactose	L-arabinose	L-tyrosine	L-arginine	D-Glucose	Acetate	palmitate	stearate	L-Threonine	L-Isoleucine	L-tryptophan	3-Methoxy-4-hydroxy-trans-cinnamate	L-Methionine	Total Weight [mg]
164.837	4.00	17.10	0.50	2.40	39.70	170000	9.20	18.00	280.00	1.10	0.50	1.10	0.06	4.00	8.00	24.00	0.90	1.20	355.00	43.00	3.00	0.90	0.80	0.80	0.20	11.00	0.50	999.961
168.737	4.04	17.15	0.54	2.43	39.65	169800	9.25	18.03	279.70	1.13	0.54	1.14	0.12	4.04	8.04	24.01	0.95	1.25	357.64	42.96	3.06	0.97	0.83	0.83	0.23	11.04	0.54	999.961
190.722	4.00	16.92	0.53	2.42	39.57	169444	9.14	17.85	277.48	1.12	0.54	1.13	0.12	4.09	7.99	23.81	0.94	1.24	361.51	42.81	3.03	0.96	0.84	0.83	0.23	10.93	0.53	999.961
192.712	3.97	16.69	0.53	2.40	39.49	169088	9.01	17.67	275.20	1.12	0.53	1.13	0.12	4.13	7.93	23.62	0.93	1.24	365.37	42.65	2.99	0.95	0.84	0.82	0.23	10.82	0.53	999.961
194.707	3.93	16.46	0.52	2.39	39.40	168732	8.89	17.49	272.93	1.11	0.53	1.12	0.12	4.17	7.87	23.42	0.92	1.24	369.25	42.49	2.95	0.93	0.83	0.82	0.23	10.71	0.53	999.961
196.707	3.90	16.23	0.52	2.38	39.32	168376	8.76	17.31	270.69	1.11	0.52	1.11	0.12	4.22	7.79	23.22	0.91	1.23	373.13	42.33	2.91	0.92	0.83	0.82	0.24	10.60	0.52	999.961
198.706	3.86	16.00	0.52	2.36	39.23	168020	8.64	17.13	268.43	1.11	0.52	1.10	0.11	4.26	7.66	23.03	0.90	1.23	377.02	42.16	2.87	0.91	0.86	0.81	0.24	10.49	0.52	999.961
200.710	3.83	15.77	0.51	2.35	39.15	167664	8.52	16.95	266.18	1.10	0.51	1.09	0.11	4.30	7.58	22.83	0.89	1.22	380.92	42.00	2.84	0.89	0.86	0.81	0.24	10.38	0.51	999.961
202.719	3.79	15.55	0.51	2.33	39.06	167308	8.39	16.77	263.93	1.10	0.51	1.08	0.11	4.34	7.51	22.63	0.88	1.22	384.82	41.83	2.80	0.88	0.87	0.80	0.24	10.27	0.51	999.961
204.732	3.75	15.32	0.50	2.32	38.97	166952	8.27	16.59	261.67	1.09	0.50	1.08	0.11	4.39	7.43	22.43	0.88	1.22	388.73	41.67	2.76	0.87	0.87	0.80	0.24	10.16	0.50	999.961
206.745	3.72	15.10	0.50	2.30	38.89	166596	8.15	16.41	259.42	1.08	0.50	1.07	0.11	4.43	7.36	22.23	0.87	1.21	392.64	41.50	2.72	0.85	0.88	0.79	0.23	10.05	0.50	999.961

Supplementary Figure 1 BioLEGO run for comparison of single-step fermentation efficiency of corn cobs by WT and RN1016 yeasts. Ten steps of gradient search for improvement of corn cob biomass composition were performed to increase ethanol production yield. The green color refers to increasing relative component weight; red color refers to decreasing relative component weight. **(a)** Maximizing minimal ethanol production and **(b)** maximizing maximal ethanol production.