

Article

Sustainable Metabolic Engineering Design Development Employing a Complex Objective Function

Reinis Muiznieks ¹, Elina Dace ¹ and Egils Stalidzans ^{1,*}

¹ Institute of Microbiology and Biotechnology, University of Latvia, Jelgavas Street 1, Riga LV1004, Latvia; lumbi@lu.lv

* Correspondence: egils.stalidzans@lu.lv

Abstract: The sustainable metabolic engineering (SME) concept was defined by Stalidzans and Dace as an approach for the selection of most sustainable metabolic engineering designs taking the economic, environmental and social components of sustainability into account. In the centre of sustainability calculations is a genome scale metabolic model that provides full balance of all incoming and outgoing metabolic fluxes at steady state. Therefore, sustainability indicators are assigned for each exchange reaction enabling calculation of sustainability features of consumption or production of each metabolite.

The further development of the SME concept depends on its implementation at the computational level to acquire applicable results – sustainable production strain designs. This study proposes for the first time a workflow and tools of SME implementation using constraint based stoichiometric modelling, genome scale metabolic models and growth coupled product synthesis approach.

For SME application demonstration purposes, a relatively simple engineering task has been carried out. The most sustainable designs have been identified using *Escherichia coli* as chassis organism, glucose as a substrate and gene deletions as metabolic engineering tool. A growth coupled production design tool has been used to reduce the variability of sustainability.

The 10 000 most sustainable designs were producing succinate as the main product with the number of deleted genes ranging from two to ten. A big number of similar designs has been identified due to the combinatorial explosion of different alternative combinations of gene deletion sets that result in interruption of the same metabolic pathways with the same impact on the metabolism.

Keywords: economic sustainability; environmental sustainability; social sustainability; genome scale metabolic model; growth coupled production

1. Introduction

The growing field of biotechnology will become a more influential fraction of the economy gradually replacing chemical (synthesis) technologies with more environmentally sustainable ones [1]. At the same time the sustainability of biotechnology itself is an important issue. Currently in many cases the limiting factor of biotechnology is the economical factor –biotechnology frequently is unable to compete with the chemical industry due to higher costs [2,3]. In other words, biotechnological solutions are less economically sustainable than chemical technologies. Knowing that there are two more components of sustainability – economic and social. Stalidzans and Dace [4] have proposed sustainable metabolic engineering (SME) as an approach for the selection of more sustainable metabolic engineering designs taking into account the economic, environmental and social components of sustainability. That way the advantages of biotechnology in environmental and social aspects over chemical technologies are taken into account and enables optimisation of all three sustainability aspects already during strain selection and design.

The assessment and design of sustainable biotechnological processes is a complex multi-criteria and multi-objective decision-making problem, not only regarding the multi-level evaluation, but also due to the significant amount of input data required for the analysis, which may come from different sources with various degrees of uncertainties [5]. Sustainability assessment is crucial in the early stages of process design as a guide for balancing economic benefits with environmental and social impacts of the operation, however, because the end design itself is still unknown, parameters necessary for sustainability assessment are uncertain and therefore estimated [6]. Although this approach comes with uncertainty, the output generated can still help in guiding sustainable design decisions [5].

The implementation of the proposed SME concept is suggested using constraint-based stoichiometric modelling [7,8] as the key element for the calculation of sustainability because

- 1) constraint based stoichiometric modelling enables analysis of an organism at genome scale providing steady state balance calculations of all incoming (substrates) and outgoing (products, biomass and by-products) metabolic fluxes and produced biomass,

- 2) the structure of the model and the metabolic flux calculation results enable assessment of different sustainability parameters using all metabolite exchange fluxes between cell and the media.

The further development of the SME concept depends on its implementation at the computational level to acquire applicable results – sustainable production strain designs. The paper of Stalidzans and Dace hints to using constraint-based modelling without an example of practical implementation of the SME concept and examples of SME-based designs.

There are several challenges to be solved implementing SME. One of them is the complexity of the objective function. Instead of a classical constraint-based modelling cases where one of the exchange reactions (usually biomass or product formation) should be maximized or minimized, in the case of SME, the objective function value must be calculated considering the impact of all exchange fluxes on the three components of sustainability. Another SME implementation issue is flux variability [9]. The flux variability is determined by the complexity and redundancy of the metabolic network and frequently leads to a variability of the exchange fluxes between organism and environment that are used in sustainability calculations. As a result, the variability of exchange fluxes causes variability of sustainability creating metabolic design ranking uncertainty.

The calculation of sustainability as an objective function can be implemented in several ways. One of the opportunities is also a promising tool of flux variability reduction by design - popular and well-developed metabolic engineering approach – growth coupled product synthesis [10–12]. Growth coupled production relies on coupling biomass production with the target product flux by deletion of some metabolic reactions to transform metabolism in a way that the fastest growing strains in the media will produce the product. Growth coupling relies on the dominance of the fastest growing strain phenotype in the media aiming to limit the metabolic behaviour solution space to minimal variability of possibly high product formation rate. Growth coupling helps also in the variability reduction at the most probable point of operation – the one with the highest growth rate.

In this study, a growth coupling algorithm was modified to implement the economic, environmental, and social sustainability calculations using a complex objective function. *Escherichia coli* is taken as a test case organism due to the relatively well known metabolism and the availability of high quality genome scale metabolic models (GSMM) [13,14]. The generated *E. coli* designs and their generation process are discussed.

2. Methods

2.1. Creation of a complex objective function for integrated sustainability score calculation

The sustainability-characterizing objective function is named integrated sustainability score (ISS) underlining the fact that ISS integrates the values of three sustainability components, each of which may be a subject of detailed calculations and analysis. ISS is introduced as a weighed sum of economic, environmental and social components of sustainability (Fig.1):

The weights can be used to bring the sustainability scores to the same unit and/or to change the priority or importance of one sustainability component over the others.

The sustainability indicators (SI) of each produced and/or consumed metabolites are stored in three SI vectors – one for each sustainability component: economic, environmental and social (Fig.1). Sustainability indicators (for example, price and environmental impact) can be assigned to a particular metabolite exchange flux. Should a part of a sustainability score be not related to a particular metabolite, it can be added for the whole design (for instance, the social impact of genetically modifying an organism cannot be related to a flux of a particular metabolite). Mathematical representation of ISS calculation and the main terms used in calculations are depicted in Figure 1.

$$\begin{aligned}
 ISS &= \sum_{i=1}^{n+k} \left(Weco \cdot v_i \cdot Sleo_i + Wenv \cdot v_i \cdot Slenv_i + Wsoc \cdot v_i \cdot SIsoc_i \right) = \\
 &= \sum_{i=1}^{n+k} v_i \cdot (Weco \cdot Sleo_i + Wenv \cdot Slenv_i + Wsoc \cdot SIsoc_i) =
 \end{aligned}$$

$\begin{pmatrix} vIn\ 01 \\ vIn\ 02 \\ vIn\ 03 \\ * \\ * \\ * \\ vIn\ n-1 \\ vIn\ n \end{pmatrix}$	$\begin{pmatrix} Weco \cdot SleoIn01 + Wenv \cdot SlenvIn01 + Wsoc \cdot SIsocIn01 \\ Weco \cdot SleoIn02 + Wenv \cdot SlenvIn02 + Wsoc \cdot SIsocIn02 \\ Weco \cdot SleoIn03 + Wenv \cdot SlenvIn03 + Wsoc \cdot SIsocIn03 \\ * \\ * \\ * \\ Weco \cdot SleoIn(n-1) + Wenv \cdot SlenvIn(n-1) + Wsoc \cdot SIsocIn(n-1) \\ Weco \cdot SleoIn\ n + Wenv \cdot SlenvIn\ n + Wsoc \cdot SIsocIn\ n \end{pmatrix}$
$= \begin{pmatrix} vOut\ 01 \\ vOut\ 02 \\ vOut\ 03 \\ * \\ * \\ * \\ vOut\ k-1 \\ vOut\ k \end{pmatrix}$	$\begin{pmatrix} Weco \cdot SleoOut01 + Wenv \cdot SlenvOut01 + Wsoc \cdot SIsocOut01 \\ Weco \cdot SleoOut02 + Wenv \cdot SlenvOut02 + Wsoc \cdot SIsocOut02 \\ Weco \cdot SleoOut03 + Wenv \cdot SlenvOut03 + Wsoc \cdot SIsocOut03 \\ * \\ * \\ * \\ Weco \cdot SleoOut\ k-1 + Wenv \cdot SlenvOut\ k-1 + Wsoc \cdot SIsocOut\ k-1 \\ Weco \cdot SleoOut\ k + Wenv \cdot SlenvOut\ k + Wsoc \cdot SIsocOut\ k \end{pmatrix}$

Figure 1. Calculation of ISS, where n – number of consumed metabolites, k – number of produced metabolites (including biomass), $Sleo_i$, $Slenv_i$ and $SIsoc_i$ – economic, environmental and social sustainability scores and $Weco$, $Wenv$ and $Wsoc$ – economic, environmental and social sustainability weights; $SSeco_i$, $SSenv_i$ and $SSsoc_i$ – economic, environmental and social sustainability scores; $WSSeco_i$, $WSSenv_i$ and $WSSsoc_i$ – weighed economic, environmental and social sustainability scores; In and Out – designations of incoming and outgoing exchange fluxes.

2.2. Calculation of sustainability indicators (SI)

To merge economic, environmental, and social impacts in one parameter, these impacts either need to share a common measurement unit or they need to be scaled with weights to create an arbitrary unit of measurement. In this study all SI were reduced to a monetary value. Using monetary value for sustainability assessment has been suggested previously [15,16]. The same measurement unit for all three sustainability components enables application of dimensionless weight coefficients.

All of them were set to “1” assigning equal importance of economic, environmental and social sustainability components in this study.

Most of the data used for Sleco and Slenv vector calculations were gathered from the BREW project [17]. The cost and environmental impact for malic acid production were gathered from the technoeconomic analysis of Bharathiraja et al. [18]. The prices for the products were gathered from the website www.chemanalyst.com during the 4th quarter of 2022. The economic impact was calculated by subtracting the production cost of a product from the price. The unit used for economic impact of each metabolite was USD/mmol. After the economic impact was found for the metabolites, they were matched to exchange reactions using Kegg database IDs [19].

A monetary value was also used for the environmental impacts as suggested by Ögmundarson et al. [16]. Although environmental impact consists of many factors, in this study only 2 of these factors were assessed – non-renewable energy use (NREU) and land use. To create the environmental impacts of the assessed metabolites, several assumptions were made. It was assumed that one disability-adjusted life year (DALY) is equivalent to 100 000 USD [15] and that lost species-years equal to 65 000 USD [20]. For environmental assessment the non-renewable energy consumption and land use from the BREW project [17] were converted to DALYs and species-years respectively. It was assumed that land use of 1 ha is equivalent to $2.5 \cdot 10^{-9}$ lost species-years [21]. These were then converted to USD as previously mentioned. The energy consumption was first converted to CO₂ equivalent assuming that 111.39 kg of CO₂ is generated to make 1 GJ of energy (<https://www.epa.gov/energy/greenhouse-gases-equivalencies-calculator-calculations-and-references>). Then the CO₂ equivalent was converted to DALYs assuming that kg of CO₂ is equal to $6.20 \cdot 10^{-7}$ DALYs [22].

Among the eight aspects of the social dimension of sustainability proposed by von Geibler and colleagues [23] five aspects do not depend on the characteristics of the used strain; therefore, they were not taken into account. The remaining three 1) health and safety, 2) employment and 3) product acceptance and social benefit are dependent on the strain design, however, only the factor of employment was quantified in this study, because the remaining two have the same values for all designs. In terms of “Health and safety”, the design of the strain can involve hazardous substances that can be avoided or reduced selecting the organism or engineering it. However, we couldn't identify any exchange metabolites used in the case study of the *E. coli* model as particularly hazardous. Hazardous substances might be involved in the downstream processes of some products; however, detailed study of the chemicals involved in these processes is out of scope for this study. The “Product acceptance and social benefit” aspect is assumed to be equal for all designs, because all of them involve genetic modifications. To assess the “Employment” aspect of different products, the number of created workplaces were taken into account assuming that every 150 000 EUR of turnover per year creates one working place [24] and that the average salary is 33 500 EUR per year (the average European salary) [25]. Therefore, the social aspect in this study was calculated as the salary gained by the working population when a product is produced. The price of the product was used to determine the turnover. This is different from the economic indicator calculation, where the possible profit is calculated by subtracting costs from the price.

The model was allowed to consume glucose as carbon source. NH₄⁺ and SO₄²⁻ were used to supply other nutrients. Glucose, NH₄⁺ and SO₄²⁻ are consumed as oppose to produced like the other metabolites, so the only variable for its economic impact is their price, which was gathered from www.chemanalyst.com as mentioned previously. In the case of NH₄⁺ and SO₄²⁻, the price of (NH₄)₂SO₄ was used, which was divided by the mol mass proportion to get the price for NH₄⁺ and SO₄²⁻. The environmental and social impacts of glucose, NH₄⁺ and SO₄²⁻ were not taken into account due to similar consumption among designs.

2.3. Sustainable metabolic engineering of *E. coli* as a case study

Escherichia coli has been used for calculations due to the availability of data and metabolic models. Genome-scale metabolic model (GSMM) iML1515 [26] has been selected for the case study. The sustainability indicators of each metabolite consumed and produced were matched with

exchange reactions of *iML1515*. ISS value (Fig.1) was used in the objective function for design evaluation and ranking. The model was set to anaerobic growth by setting the oxygen exchange reaction bounds to 0. All the other exchange reactions were left unchanged; therefore, glucose was used as the substrate at a consumption rate of 10 mmol/gDW/h.

SME approach has been implemented using COBRA toolbox [27] for calculations, while Paint4net [28] and IMFLer [29] were used for visualisation. The optimization was run with *iML1515*, allowing the deletions of genes that were determined as non-essential from Baba et al. [30].

2.4. Design development with genetic algorithms

To create designs using sustainability as objective function, *optGene* [31] was used as a growth-coupling framework. Genetic algorithms use populations of individuals, which can mutate and evolve, to maximize a score which is derived from a fitness function [32]. In the case of *optGene*, each individual is a design that consists of either reactions or genes to be deleted. The default fitness function of *optGene* calculates the minimal production rate of a desired reaction at the maximum growth rate of a design. To create the designs in this publication, *optGene* was modified to work with the ISS as an objective function. The obtained objective function was introduced into the *optGeneFitnessTilt.m* function as a global variable, where it was used to assess the fitness score of individual designs from the population. The fitness score was calculated using the minimal sustainability values at the minimal and maximal growth-rates of a design. To limit the use of unnecessary knockouts, the number of knockouts was used to negatively impact the fitness score. The fitness score of an individual was derived from the equation (1),

$$\begin{aligned} \text{Fitness score} \\ = -30 - ISS_{AtMaxGR} - 0.5 \cdot ISS_{AtZeroGR} + 0.1 \\ \cdot nrOfKO \quad (1) \end{aligned}$$

where $ISS_{AtZeroGR}$ – the integrated sustainability score at minimal growth-rate, $ISS_{AtMaxGR}$ – the integrated sustainability score at maximal growth-rate and $nrOfKO$ – number of knockouts in a design.

The ISS values used in the *optGene* fitness function are visualized in Figure 2. It should be noted that *optGene* is designed to minimize the fitness score, therefore, the negatives of the ISS are used. The subtraction of 30 from the fitness score is used to make all the individuals from the population have negative fitness score, because the minimal sustainability of a design can be a negative value, therefore, making the fitness score positive. It was found that when positive fitness scores are allowed, the genetic algorithm task stagnates, as *optGene* assigns a fitness score of 0 to designs which can't obtain growth. This in turn can make the fittest individuals at the start of the optimization process the ones who are infeasible. Because mostly the fittest individuals are used to create the next populations, the process stagnates. *optGene* was also modified to return the hashtable of all the generated designs, so that not only the fittest design is accessible after the optimization, but also other less fit designs can be analyzed.

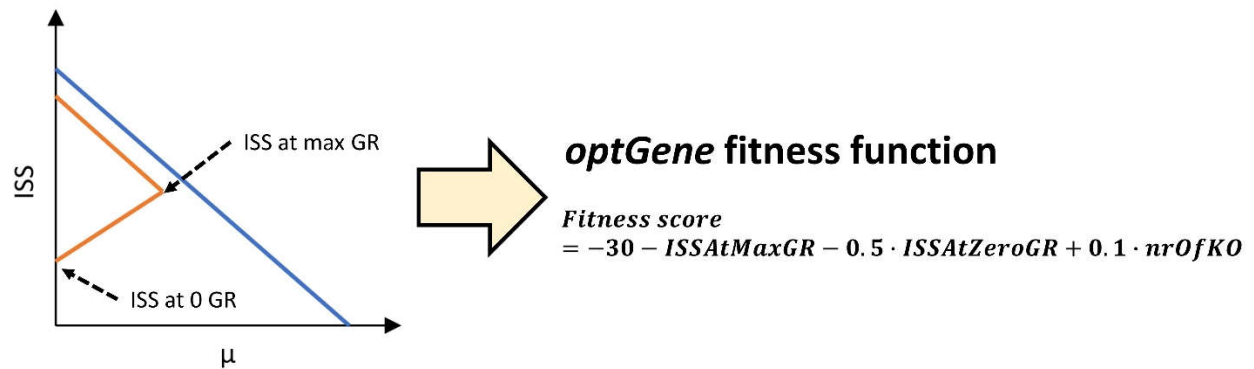


Figure 2. The points from the envelope used in the *optGene* fitness function. Blue and orange lines represent wild-type and metabolically engineered strain envelopes, respectively. ISS – integrated sustainability score, μ – growth-rate, GR – growth-rate, KO – gene knockouts.

In the SME approach the objective function is ISS instead of product or biomass. Therefore, selection of target products is an outcome of optimization determined by the way a product can improve the three components of sustainability.

A maximum of 7 gene deletions were allowed in the *E. coli* (section 2.2.) use case. The *optGene* task was run until stagnation was achieved, which took around 2 hours. The SME implementation code in MATLAB (R2020a) is available online at https://github.com/lv-csbg/SME_code.

3. Results

3.1. Complex objective function for sustainable metabolic engineering

To assess sustainability using genome-scale metabolic models (GSMMs), the flux of each exchange reaction can be multiplied by a corresponding value, which signifies the sustainability when the metabolite is either produced or consumed. This value should come from the economic, environmental, and social indicators from the production or consumption of the metabolite and is called sustainability indicator (SI) [4]. The objective function of a GSMM (the *c* vector of a COBRA model structure) holds a value for each reaction in the model, which during optimization (COBRA *optimizeCbModel.m* function) is multiplied by the corresponding reactions' flux and summed together. Thus, if the objective function holds coefficients for the sustainability impact of exchange reactions, sustainability can be assessed.

Sustainability indicator values for different metabolite products were created as described in the Methods section. Table 1 shows the SI values, the economic, environmental and social indicators that make up this coefficient, and the corresponding metabolites. The environmental and economic impacts are mostly made up from the downstream processing of the exchange products. The highest sustainability impact of a product as seen in Table 1 is for malic acid and the lowest for lactic acid. The only substrate in Table 1 is glucose, which also has a low SI value that is made up only by its economic impact. The SI of glucose is positive, however, when glucose is consumed as a substrate, its exchange reaction has a negative value. Therefore, during optimization with the sustainability objective function glucose has a negative impact on the optimization value. The opposite can be said about CO₂, which has a negative SI, made up from its environmental impact.

The sustainability indicator values for the metabolites from Table 1 were matched with their corresponding exchange reactions to calculate ISS (Fig.1). The sustainability objective function was used with the COBRA model structure and optimized with the *optimizeCbModel.m* COBRA function to obtain the ISS. The objective value returned represents sustainability in USD per gram of biomass per hour. By changing the bounds of the biomass reaction, values for the minimum and maximum sustainability at specific growth rates can be obtained. Thus, a *sustainability envelope* can be obtained instead of the traditional production envelope (Fig.2). Furthermore, this sustainability envelope can

be used to rank growth-coupled designs by the features of the production envelope. Designs are ranked according to the envelope fitness function (Equation 1, Fig.2).

Table 1. Metabolites with their corresponding economic, environmental, and social SI and their sum (integrated sustainability indicator).

Metabolite	Economic SI (1e-4 USD/mmol)	Environmental SI (1e-4 USD/mmol)	Social SI (1e-4 USD/mmol)	Integrated SI (1e-4 USD/mmol)
Acrylic acid	1.561	-0.022	0.486	2.025
Adipic acid	2.311	-0.864	1.217	2.664
Citric acid	2.299	-0.977	0.849	2.171
CO ₂	0	-0.027	0	-0.027
Ethanol	0.382	0.044	0.121	0.547
Ethylene	0.078	0.058	0.063	0.198
Glucose	1.042	0.000	0.233	1.275
Lactic acid	0.434	-0.098	0.249	0.586
Malic acid	6.048	-0.153	1.480	7.376
Succinic acid	3.359	-0.366	0.957	3.950
NH ₄ ⁺	0.017	0	0	0.017
SO ₄ ²⁻	0.235	0	0	0.235

3.2. Sustainable metabolic engineering designs

By using the created sustainability objective function together with the genetic algorithm framework *optGene* metabolic engineering designs were obtained. The advantage of using *optGene* instead of other growth-coupling software is the ability to use different properties of the GSMM designs in the fitness function. In this study, the fitness function of *optGene* was modified according to equation (1) (see Methods). A maximum of 7 knockouts were allowed. The fitness score includes a penalty for the number of gene knockouts so that designs with marginally better or the same sustainability scores, which require more gene knockouts, are less fit.

Thousands of designs have been generated and ranked according to equation (1). All the top designs produce succinate as the main product. The first 9000 designs are very similar by their sustainability envelope characterizing parameters (see the first 1000 designs in **Supplementary Table S1**). The high number of similar alternatives arises because of the combinatorial explosion of different deleted gene combinations that result in interruption of the same metabolic reaction or pathway yielding the same sustainability parameters. The high number of alternatives enables selection of the most convenient gene deletion set. Starting from rank 9000 the designs start to become slightly different (see the 9000- 10000 ranked designs in **Supplementary Table S2**). The hashtable containing all the evaluated designs and their fitness scores is also available as **Supplementary File S3**). In Table 2, designs ranked 1, 9025 and 9031 are compared with the *E. coli* wild-type.

Table 2. Parameters of the designs #1, #9025, #9032 and wild-type.

Design	#1	#9025	#9031	Wild-type
Fitness score	-59.065	-50.4687	-49.623	-15.133
Main product	Succinate	Succinate	Succinate	Ethanol
Main production flux at max GR(mmol/gDW/h)	10.806	10.391	10.391	8.746
Max GR (1/h)	0.069	0.071	0.071	0.158
Number of gene knockouts	3	7	2	0
Minimum ISS at maximum GR (1e-4*USD/gDW/h)	31.702	30.372	28.413	-7.805
Minimum ISS at 0 GR (1e-4*USD/gDW/h)	-5.752	-6.370	-13.581	-14.125
Economic SS at maximal GR (1e-4*USD/gDW/h)	27.072	25.909	24.469	-6.943
Environmental SS at maximal GR (1e-4*USD/gDW/h)	-3.759	-3.603	-3.664	0.363
Social SS at maximal GR (1e-4*USD/gDW/h)	8.389	8.066	7.609	-1.225
Gene knockouts	b2415, b3737, b1380	b0894, b4025, b2988, b2428, b3063, b3736, b1380	b4025, b3733	

The sustainability envelopes of designs #1, #9025, #9031 and wild-type are shown in Figure 3. The envelopes of the designs are quite similar and differ greatly from the wild-type. The ISS at maximal and 0 growth-rate of design #1 are higher than the same values for designs #9025 and #9031. Design #9031 uses only 2 gene deletions but designs #1 and #9025 use 3 and 7 gene deletions respectively. The benefits of the designs should be weighed against the difficulties of implementation. These similar envelopes suggest that the gene deletion cost in the fitness function (Equation 1) could be increased if the extra deletions seem to offer results which are not worth the implementation.

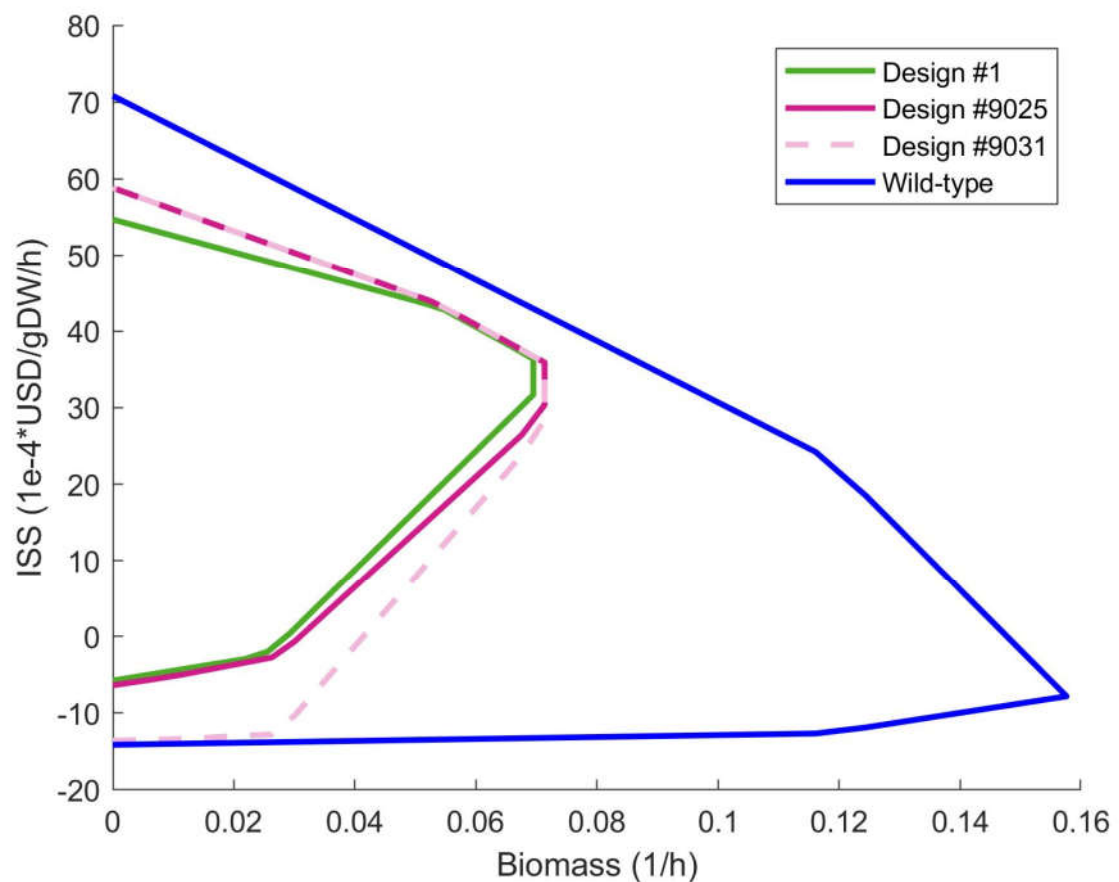


Figure 3. Sustainability envelopes of designs ranked #1, #9025, #9031 and wild-type strain. ISS – integrated sustainability score.

Although succinate is the main contributor of the ISS for the 3 designs, the ISS is made up from exchange fluxes that have reaction sustainability coefficients assigned to them. In Table 3. the exchange reactions active and their impact on the ISS are shown for design #1 at maximal growth. Design #1 produces not only succinate but also ethanol, although at a lower rate. Extracting both ethanol and succinate from the medium could be considered if possible.

Table 3. Exchange reactions at minimal ISS and maximal growth for design #1. ISS- integrated sustainability score, SS – sustainability score.

Exchange reaction	Reacti sustainability coefficient (1e-4 USD/mmol)	Reaction flux (mmol/gDW/h)	ISS (1e-4 USD/gDW)	Economic SS (1e-4 USD/gDW)	Environmenta l SS (1e-4 USD/gDW)	Social SS (1e-4 USD/gDW)
CO ₂	-0.03	-2.09	0.06	0.00	0.06	0.00
Succinate	3.95	10.81	42.68	36.30	-3.95	10.34
Glucose	1.27	-10.00	-12.75	-10.42	0.00	-2.33
SO ₄ ²⁻	0.29	-0.02	-0.01	0.00	0.00	0.00
NH ₄ ⁺	0.02	-0.75	-0.02	-0.01	0.00	0.00
Ethanol	0.55	3.17	1.73	1.21	0.14	0.38

4. Discussion

A method for assessing and optimizing sustainability at early strain design stages according to sustainable metabolic engineering (SME) concept using GSMM is demonstrated. The use of a complex objective function together with genetic algorithms within framework of growth coupled

production allows optimization of economic, environmental and social sustainability during *in silico* metabolic engineering design development. Generally, the SME mission is finding the best combination of available variables: organism, substrate, product and metabolic engineering interventions. As a result, mentioned variables, that are usually pre-defined during design of biotechnological production, become instruments of sustainability improvement in the case of SME.

SME can be useful also when choosing the best production process for a specific industrial setting, substrate, organism or product. That means, some SME task setting variables are fixed. To implement that, the sustainability related features of fixed variables should be artificially improved by correction of corresponding SI values. The list of candidate organisms can be limited by the number of models made available for the algorithm.

In the current study the main aim was to demonstrate practically the implementation of SME in a simplified task setting. Therefore, only one organism (*E. coli.*) with a single substrate (glucose) was used with a limited number of exchange metabolites. For simplicity insertions of reactions were not considered. Even this simple task setting resulted in a large number of designs indicating the computational challenge if more degrees of freedom are enabled. The proposed approach has been applied in another study for sustainable design development to utilize amino acids of spent microbial biomass [33].

The three pillars of sustainability – economic, environmental, and social impact – were assessed for 7 metabolites in a simplified way. Generally simplified values for the first screening of designs by SME tools may be sufficient to identify the designs that stand out among the others by great sustainability features. In the next stage more accurate information about the key metabolites could be collected to assess smaller number of designs with higher accuracy. For example, the environmental impact was assessed using only non-renewable energy use (NREU) and land use, which were then converted to monetary value. Life cycle analysis (LCA) could be used to better assess the environmental impact, by including more impact categories, such as eutrophication, ecotoxicity and acidification [34]. However, to gather more accurate sustainability coefficients for each metabolite, more specific information about the production process would be needed. The available information about different biotechnologically produced products varies in quality. For example, LCA information about lactic and succinic acid production is much more available than for 1,5-pentanediamine production while the latter product itself is much more toxic to human health [34,35]. This again shows the complexity of early sustainability assessment during production design. The quantification of more social sustainability indicators than salary is not implemented during this study, because two other strain design related parameters were equal for all designs: we did not use any dangerous metabolite in terms of health and safety and all designs contained genetic manipulations. Generally, SME is intended for screening of design ideas and the accuracy of SI data at SME stage is not critical, because a small number of most interesting designs should undergo more careful analysis using LCA, techno-economic analysis or other approaches.

Keeping all of this in mind, the point of this study is to show the framework of generating sustainable metabolic engineering designs, not the sustainability assessment values themselves. In other words, we demonstrate the calculation workflow where data is one of the inputs. Better quality input data will yield designs with more accuracy and reliability.

The optimization of *E. coli* showed that the sustainability indicators of products and the sustainability envelope of the wild-type organism doesn't predict the most sustainable product. The highest SI values in this study were for malic acid production. Malic acid was also found to be produced at the maximum sustainability points at different growth-rates for wild-type *E. coli*. However, after optimization no design was found with malic acid as its main product in the top 1000 designs. Although the best designs might change when the SI values are altered, our findings suggest that it is not trivial to pick out which product will assure the best sustainability of design. In this study a framework of early-stage sustainability assessment was shown, however, more research is needed to better assess the sustainability indicators of each product. Detailed and standardized LCA models are needed to properly assess the sustainability using a GSMM. A future perspective could be to use the byproduct production as input values for LCA analysis instead of using values from

already existing processes. This would mean that the purification and waste management, for example, could be quantified using GSMM.

Supplementary Materials: Table S1: Top 1000 ranked designs; Table S2: 9000-10000 ranked designs; File S3: Matlab object with all evaluated designs and their fitness scores.

Author Contributions: Conceptualization, R.M., and E.S.; methodology, E.S. and R.M.; software, R.M.; investigation, E.D. and R.M.; resources, E.S.; data curation, R.M. and E.D.; writing—original draft preparation, E.S. and R.M.; writing—review and editing, R.M, E.S. and E.D.; visualization, R.M.; supervision, E.S.; project administration, E.S.; funding acquisition, E.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported by the European Regional Development Fund within the project No. 1.1.1.1/20/A/137 “genome scale metabolic modelling linked bioreactor control system (GenCon)” and University of Latvia project “Optimization of biotechnological processes for effective utilization of renewable resources” (Nr. Y5-AZ20-ZF-N-270).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bennich, T.; Belyazid, S. The Route to Sustainability—Prospects and Challenges of the Bio-Based Economy. *Sustainability* **2017**, *9*, 887, doi:10.3390/su9060887.
2. Chen, T.-L.; Kim, H.; Pan, S.-Y.; Tseng, P.-C.; Lin, Y.-P.; Chiang, P.-C. Implementation of Green Chemistry Principles in Circular Economy System towards Sustainable Development Goals: Challenges and Perspectives. *Science of The Total Environment* **2020**, *716*, 136998, doi:10.1016/j.scitotenv.2020.136998.
3. Ramos, J.L.; Duque, E. Twenty-First-Century Chemical Odyssey: Fuels versus Commodities and Cell Factories versus Chemical Plants. *Microb Biotechnol* **2019**, *12*, 200–209, doi:10.1111/1751-7915.13379.
4. Stalidzans, E.; Dace, E. Sustainable Metabolic Engineering for Sustainability Optimisation of Industrial Biotechnology. *Comput Struct Biotechnol J* **2021**, *19*, 4770–4776, doi:10.1016/j.csbj.2021.08.034.
5. Gargalo, C.L.; Sin, G. Sustainable Process Design under Uncertainty Analysis: Targeting Environmental Indicators. *Computer Aided Chemical Engineering* **2015**, *37*, 2579–2584, doi:10.1016/B978-0-444-63576-1.50124-2.
6. Gargalo, C.L.; Cheali, P.; Posada, J.A.; Carvalho, A.; Gernaey, K. v.; Sin, G. Assessing the Environmental Sustainability of Early Stage Design for Bioprocesses under Uncertainties: An Analysis of Glycerol Bioconversion. *J Clean Prod* **2016**, *139*, 1245–1260, doi:10.1016/j.jclepro.2016.08.156.
7. O'Brien, E.J.; Monk, J.M.; Palsson, B.O. Using Genome-Scale Models to Predict Biological Capabilities. *Cell* **2015**, *161*, 971–987, doi:10.1016/j.cell.2015.05.019.
8. Price, N.D.; Reed, J.L.; Palsson, B.Ø. Genome-Scale Models of Microbial Cells: Evaluating the Consequences of Constraints. *Nat Rev Microbiol* **2004**, *2*, 886–897, doi:10.1038/nrmicro1023.
9. Gudmundsson, S.; Thiele, I. Computationally Efficient Flux Variability Analysis. *BMC Bioinformatics* **2010**, *11*, 489, doi:10.1186/1471-2105-11-489.
10. Motamedian, E.; Berzins, K.; Muiznieks, R.; Stalidzans, E. OptEnvelope: A Target Point Guided Method for Growth-Coupled Production Using Knockouts. *bioRxiv* **2023**.
11. von Kamp, A.; Klamt, S. Growth-Coupled Overproduction Is Feasible for Almost All Metabolites in Five Major Production Organisms. *Nat Commun* **2017**, *8*, 15956, doi:10.1038/ncomms15956.
12. Klamt, S.; Mahadevan, R. On the Feasibility of Growth-Coupled Product Synthesis in Microbial Strains. *Metab Eng* **2015**, *30*, 166–178, doi:10.1016/j.ymben.2015.05.006.

13. Oberhardt, M. a; Palsson, B.O.; Papin, J. a Applications of Genome-Scale Metabolic Reconstructions. *Mol Syst Biol* **2009**, *5*, 320, doi:10.1038/msb.2009.77.
14. Fell, D.A.; Poolman, M.G.; Gevorgyan, A. Building and Analysing Genome-Scale Metabolic Models. *Biochem Soc Trans* **2010**, *38*, 1197–1201, doi:10.1042/BST0381197.
15. Dong, Y.; Hauschild, M.; Sørup, H.; Rousselet, R.; Fantke, P. Evaluating the Monetary Values of Greenhouse Gases Emissions in Life Cycle Impact Assessment. *J Clean Prod* **2019**, *209*, 538–549, doi:10.1016/j.jclepro.2018.10.205.
16. Ögmundarson, Ó.; Sukumara, S.; Herrgård, M.J.; Fantke, P. Combining Environmental and Economic Performance for Bioprocess Optimization. *Trends Biotechnol* **2020**, *38*, 1203–1214, doi:10.1016/j.tibtech.2020.04.011.
17. BREW, G. agreement I.G.-C.-2002-00014 *Medium and Long-Term Opportunities and Risks of the Biotechnological Production of Bulk Chemicals from Renewable Resources*; 2006;
18. Bharathiraja, B.; Jayamuthunagai, J.; Sreejith, R.; Iyyappan, J.; Praveenkumar, R. Techno Economic Analysis of Malic Acid Production Using Crude Glycerol Derived from Waste Cooking Oil. *Bioresour Technol* **2022**, *351*, 126956, doi:10.1016/j.biortech.2022.126956.
19. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Res* **2023**, *51*, D587–D592, doi:10.1093/nar/gkac963.
20. Weidema, B.P. Using the Budget Constraint to Monetise Impact Assessment Results. *Ecological Economics* **2009**, *68*, 1591–1598, doi:10.1016/j.ecolecon.2008.01.019.
21. Chaudhary, A.; Verones, F.; De Baan, L.; Pfister, S.; Hellweg, S. 11. Land Stress: Potential Species Loss from Land Use (Global; PSSRg). In; 2016.
22. Eckelman, M.J.; Sherman, J.D. Estimated Global Disease Burden From US Health Care Sector Greenhouse Gas Emissions. *Am J Public Health* **2018**, *108*, S120–S122, doi:10.2105/AJPH.2017.303846.
23. von Geibler, J.; Liedtke, C.; Wallbaum, H.; Schaller, S. Accounting for the Social Dimension of Sustainability: Experiences from the Biotechnology Industry. *Bus Strategy Environ* **2006**, *15*, 334–346, doi:10.1002/bse.540.
24. Rafiaani, P.; Kuppens, T.; Dael, M. Van; Azadi, H.; Lebailly, P.; Passel, S. Van Social Sustainability Assessments in the Biobased Economy: Towards a Systemic Approach. *Renewable and Sustainable Energy Reviews* **2018**, *82*, 1839–1853, doi:10.1016/j.rser.2017.06.118.
25. Eurostat New Indicator on Annual Average Salaries in the EU.
26. Monk, J.M.; Lloyd, C.J.; Brunk, E.; Mih, N.; Sastry, A.; King, Z.; Takeuchi, R.; Nomura, W.; Zhang, Z.; Mori, H.; et al. IML1515, a Knowledgebase That Computes Escherichia Coli Traits. *Nat Biotechnol* **2017**, *35*, 904–908, doi:10.1038/nbt.3956.
27. Heirendt, L.; Arreckx, S.; Pfau, T.; Mendoza, S.N.; Richelle, A.; Heinken, A.; Haraldsdóttir, H.S.; Wachowiak, J.; Keating, S.M.; Vlasov, V.; et al. Creation and Analysis of Biochemical Constraint-Based Models Using the COBRA Toolbox v.3.0. *Nat Protoc* **2019**, *14*, 639–702, doi:10.1038/s41596-018-0098-2.
28. Kostromins, A.; Stalidzans, E. Paint4Net: COBRA Toolbox Extension for Visualization of Stoichiometric Models of Metabolism. *Biosystems* **2012**, *109*, 233–239, doi:10.1016/j.biosystems.2012.03.002.
29. Petrovs, R.; Stalidzans, E.; Pentjuss, A. IMFLer: A Web Application for Interactive Metabolic Flux Analysis and Visualization. *Journal of Computational Biology* **2021**, *28*, 1–12, doi:10.1089/cmb.2021.0056.
30. Baba, T.; Ara, T.; Hasegawa, M.; Takai, Y.; Okumura, Y.; Baba, M.; Datsenko, K.A.; Tomita, M.; Wanner, B.L.; Mori, H. Construction of Escherichia Coli K-12 In-frame, Single-gene Knockout Mutants: The Keio Collection. *Mol Syst Biol* **2006**, *2*, doi:10.1038/msb4100050.

31. Patil, K.R.; Rocha, I.; Förster, J.; Nielsen, J. Evolutionary Programming as a Platform for in Silico Metabolic Engineering. *BMC Bioinformatics* **2005**, *6*, 308, doi:10.1186/1471-2105-6-308.
32. Forrest, S. Genetic Algorithms. *ACM Comput Surv* **1996**, *28*, 77–80, doi:10.1145/234313.234350.
33. Stikane, A.; Baumanis, M.R.; Muiznieks, R.; Stalidzans, E. Impact of Waste as a Substrate on Biomass Formation, and Optimisation of Spent Microbial Biomass Re-Use by Sustainable Metabolic Engineering. *Fermentation* (submitted) **2023**.
34. Ögmundarson, Ó.; Herrgård, M.J.; Forster, J.; Hauschild, M.Z.; Fantke, P. Addressing Environmental Sustainability of Biochemicals. *Nat Sustain* **2020**, *3*, 167–174, doi:10.1038/s41893-019-0442-8.
35. Til, H.P.; Falke, H.E.; Prinsen, M.K.; Willems, M.I. Acute and Subacute Toxicity of Tyramine, Spermidine, Spermine, Putrescine and Cadaverine in Rats. *Food and Chemical Toxicology* **1997**, *35*, 337–348, doi:10.1016/S0278-6915(97)00121-X.