Die Grenzen der Chemie neu ausloten? It takes #HumanChemistry

Wir suchen kreative Chemikerinnen und Chemiker, die mit uns gemeinsam neue Wege gehen wollen – mit Fachwissen, Unternehmertum und Kreativität für innovative Lösungen. Informieren Sie sich unter:

evonik.de/karriere

EVONIK
Leading Beyond Chemistry

Received: 4 May 2021 | Revised: 10 August 2021 | Accepted: 11 August 2021

DOI: 10.1002/biot.202100236

Biotechnology Journal

RESEARCH ARTICLE

Systematizing the different notions of growth-coupled product synthesis and a single framework for computing corresponding strain designs

Philipp Schneider[1] | **Radhakrishnan Mahadevan**[2] | **Steffen Klamt**[1]

[1] Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

[2] Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada

Correspondence
Steffen Klamt, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, Magdeburg, D-39106, Germany.
Email: klamt@mpi-magdeburg.mpg.de

Funding information
H2020 European Research Council, Grant/Award Number: 721176

Abstract

A widely used design principle for metabolic engineering of microorganisms aims to introduce interventions that enforce growth-coupled product synthesis such that the product of interest becomes a (mandatory) by-product of growth. However, different variants and partially contradicting notions of growth-coupled production (GCP) exist. Herein, we propose an ontology for the different degrees of GCP and clarify their relationships. Ordered by coupling degree, we distinguish four major classes: potentially, weakly, and directionally growth-coupled production (pGCP, wGCP, dGCP) as well as substrate-uptake coupled production (SUCP). We then extend the framework of Minimal Cut Sets (MCS), previously used to compute dGCP and SUCP strain designs, to allow inclusion of implicit optimality constraints, a feature required to compute pGCP and wGCP designs. This extension closes the gap between MCS-based and bilevel-based strain design approaches and enables computation (and comparison) of designs for all GCP classes within a single framework. By computing GCP strain designs for a range of products, we illustrate the hierarchical relationships between the different coupling degrees. We find that feasibility of coupling is not affected by the chosen GCP degree and that strongest coupling (SUCP) requires often only one or two more interventions than wGCP and dGCP. Finally, we show that the principle of coupling can be generalized to couple product synthesis with other cellular functions than growth, for example, with net ATP formation. This work provides important theoretical results and algorithmic developments and a unified terminology for computational strain design based on GCP.

KEYWORDS
computational strain design, constraint-based modeling, metabolic engineering

Abbreviations: ACP, ATP-coupled production; FBA, flux balance analysis; GCP, growth-coupled production; MCS, minimal cut set; MILP, mixed integer linear programming; NGAM, non-growth associated maintenance; pACP, potentially ATP-coupled production; PE, production envelope; pGCP, potentially growth-coupled production; PYS, production yield space; dACP, directionally ATP-coupled production; dGCP, directionally growth-coupled production; SUCP, substrate-uptake coupled production; wACP, weakly ATP-coupled production; wGCP, weakly growth-coupled production

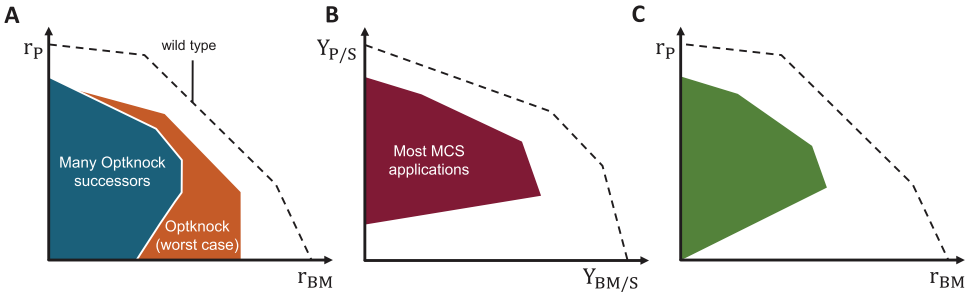

FIGURE 1 Characteristic production envelopes or yield spaces of different computational methods for growth-coupled strain design. (A) Typical production envelopes of biased (bilevel) optimization techniques. In the worst case, the original method OptKnock (orange) may contain flux vectors with no product synthesis at maximal growth rate, which is avoided by successors of OptKnock (blue). (B) Typical yield space of a strain design computed with MCS, demanding a minimum product yield for all flux states. (C) Production envelope of a strain design with a fixed minimum ratio of product synthesis and growth rate. r_{BM}: growth rate; r_P: product synthesis rate; $Y_{P/S}$: product yield; $Y_{BM/S}$: biomass yield

1 | INTRODUCTION

Metabolic engineering aims to harness and improve the production capabilities of microbial organisms and to prime them for the use in bioproduction processes. Wild type microbes usually tune their metabolic activities to grow at the highest rate possible. This goal conflicts with the engineering goal of increased product synthesis. It is therefore often desirable to hardwire a dependence between growth and product synthesis through suitable gene knockouts or overexpressions and make the product of interest a mandatory by-product of growth.[1–8] Strains in which growth is coupled with product synthesis exhibit exponential formation of both biomass and product. The potential of this powerful design principle could be demonstrated in several successful metabolic engineering studies.[4,9–15] Finding the metabolic interventions that enforce growth-coupled product synthesis is also a common goal of many computational strain design methods.[16,17] Recently, it has been shown that, in principle, almost all small metabolites in five major production organisms can be coupled with growth by suitable knockout strategies.[8]

Different degrees or types of growth-coupled production (GCP) have been used in the strain design algorithms developed in the past. The first computational method for growth-coupled strain design was OptKnock.[1] One important motivation of the OptKnock method was that growth-coupled strain design can be combined with adaptive laboratory evolution strategies to evolve constructed strains towards their growth-maximal phenotype and, thus, also towards maximal product synthesis.[18] With that, it was considered to be sufficient to demand GCP only for the growth-optimal flux states, while flux distributions with sub-optimal growth can be disregarded. OptKnock implemented this idea through a nested (bilevel) maximization approach, which generates strain designs with a potential for product synthesis at maximal growth. However, OptKnock strain design does not *guarantee* the exploitation of that production potential and solutions may exist where maximal growth is also possible without any product synthesis due to the presence of alternate optimal solutions (orange area in Figure 1A). Such properties can be best analyzed in the production envelope ((PE), also called growth-product phase plane or biomass-vs.-product tradeoff plot), a projection of all steady-state flux vectors of the (designed) metabolic network on their growth rate (r_{BM}) and product synthesis rate (r_P).[16,19] OptKnock served as a starting point for the development of various related methods.[2,3,6,20–25] Common to many successors of OptKnock is that they explicitly enforce some minimum product synthesis at all growth-maximal flux states while GCP for suboptimal growth rates is, as in OptKnock, usually not demanded (Figure 1A; blue area). Hence, the coupling of growth with product synthesis is in most cases still only a local property and does not include the entirety of possible growth rates. Because these bilevel optimization approaches assume that maximal growth rates are attained by the mutant strains (possibly after adaptive laboratory evolution), they are sometimes called "biased" strain design methods.[16]

Alternative approaches developed by Trinh and co-workers[10,24,26,27] as well as the framework of Minimal Cut Set (MCS),[28,29] sought to establish a more rigid coupling (also called 'strong coupling') by enforcing a certain minimum product yield in all (non-zero) flux states. The two-dimensional biomass-product-yield spaces of such designed strains have typically a shape as shown in Figure 1B. In contrast to the PE, the biomass-product-yield space (in the following also called production yield space (PYS)) maps all feasible steady state flux solutions of the strain to biomass yield ($Y_{BM/S}$) and product yield ($Y_{P/S}$). Although, this stronger notion of GCP is commonly used in the context of MCS[28,30] or for the construction of modular cells,[24,31] it seems to demand more than necessary because production is also enforced in metabolic states without growth. The relationships between computed bilevel and the strongly growth-coupled strain designs have not been studied so far.

An intermediate variant of GCP with a coupling strength lying between the classical bilevel and the strong coupling approach can be defined by demanding a minimum ratio of product synthesis rate and growth rate. The flux spaces of the corresponding strain designs appear as areas in the PE that are entirely above a line with a certain slope (Figure 1C). This definition of GCP was proposed in few theoretical works, partially under different names[6,32,33]; however, so far, it has not been used in any application and deserves further attention.

To characterize the different types and strengths of GCP approaches, many studies introduced attributes such as weak,[6,7,24,32]

maximal,[1] holistic,[32] tight,[26,27] directional,[6] partial and full,[16] or strong coupling.[7,24,32] However, this resulted in partially inconsistent or imprecise terminology in the literature, caused by parallel developments or by employing different mathematical approaches and their respective reference points (e.g., PE vs. yield space, bilevel optimization vs. MCS etc.). Moreover, some definitions are not conclusive enough to characterize or/and distinguish different types of coupling. For example, the PE shown in Figure 1C may also be exhibited by a strain that has the stronger coupling degree depicted in Figure 1B, hence, Figure 1C alone is not sufficient to distinguish these two coupling types.

One goal of this study is to systematize and unify the different definitions of (the degree of) GCP to obtain one consistent ontology. We distinguish four major types of coupling based on unambiguous definitions and show that the MCS framework can be used to compute strain designs for all four coupling degrees. In particular, we will present extensions of the MCS approach that now also enable the computation of classical bilevel strain designs via MCS by incorporating implicit constraints for growth-rate optimality. We then compare strain designs computed for 12 (native and heterologous) products in *Escherichia coli* for all four coupling types regarding the number of required interventions and computation time. Finally, we also discuss a generalization of GCP towards ATP-coupled product synthesis, which is relevant for other metabolic engineering strategies.

2 | METHODS

2.1 | Definition of four coupling degrees

The key property of GCP is the dependence of cellular growth on the synthesis of a product of interest. We propose four different degrees of this relationship, which are summarized in Figure 2 and detailed below.

The first coupling degree is called potentially growth-coupled product synthesis (pGCP) and occurs when there is potential for product synthesis at growth-maximal flux states. Using @ as abbreviation for "at", we express this condition by

$$r_P^{max@r_{BM}^{max}} > 0. \quad (1)$$

pGCP formulates the mildest condition of all coupling degrees, and OptKnock is the most popular method for computing pGCP strains designs. In contrast to the other three coupling degrees to follow, pGCP does not ensure a strict dependency between growth and production. However, it indicates that product synthesis does not oppose to the biological objective of growth maximization.

Clearly, definition (1) is only meaningful if there is at least one flux vector with a non-zero growth rate, and in the following we assume that this is always fulfilled in the system under study. Figure 2 provides examples of interventions that induce the respective coupling degrees in a given example network. For this network, we assume that substrate uptake is the only reaction with an upper bound for its flux and that P is our desired product. Biomass is here represented by an essential biomass precursor BM. In this network, pGCP can be induced through a single knockout: removing the reaction from A to BM ensures that the pathway with BM-optimal but product-free operation is inactive. Two alternative pathways with maximal growth rates remain, one has P and the other Q as by-product. It is possible to shift metabolic flux between both pathways and produce the product P or Q, without a decrease in the maximum attainable growth rate.

Generally, whether a certain growth-coupling degree is prevalent in a metabolic network or not can be tested through simple flux optimizations (flux balance analysis (FBA)[34]) and yield optimizations[33] or, alternatively, inferred graphically from the PE and PYS as shown in Figure 2. Regarding pGCP, its presence can be tested via FBA by first maximizing the growth rate followed by a second maximization of the production rate with the growth rate constrained at its maximum. pGCP is present when the production rate in the second optimization takes a strictly positive value. In the PE, for pGCP it is required that there is at least one point with maximal growth rate that lies above the x-axis. In our example, we have a vertical line at maximal growth starting from the x-axis (zero production of P) (Figure 2), indicating that maximal growth may coincide with simultaneous production but that alternative routes with maximal BM synthesis but without production of P do exist. Importantly, despite the fact that, in the example, the PYS for pGCP shows the same characteristic shape as the PE, potential coupling cannot generally be deduced from the PYS because flux states with maximal growth rate may not exhibit maximal biomass yield (which is at the rightmost position in the PYS).[33] A mismatch occurs when the capacity of pathways with optimal biomass yield is reached due to inhomogeneous constraints (such as a maximal oxygen uptake rate) and suboptimal pathways become active to attain the maximal growth rate. This happens, for example, when overflow metabolism occurs as a consequence of proteome allocation constraints.[35]

Next, if the dependence of growth on production is present in all flux states with maximal growth rate, then we call it weakly growth-coupled production (wGCP). We express this condition with

$$r_P^{min@r_{BM}^{max}} > 0. \quad (2)$$

An example of a wGCP design strategy in the toy network is also shown in Figure 2. After the additional knockout of the reaction converting E to F, the only pathway left to offer the highest growth rate would run via E to BM, and this pathway produces P as byproduct as desired. The presence of wGCP in a network can be tested via FBA by first maximizing the growth rate and subsequently minimizing the production rate, with the growth rate constrained at its maximum. If the latter minimization returns a strictly positive value, wGCP is present. Alternatively, wGCP can also be directly inferred from the PE: there must be an edge or a corner at maximal growth that does not touch the x-axis). As for pGCP, wGCP cannot generally be deduced from the PYS. Theoretically, as an alternative definition for weak coupling, one could demand some minimum product yield at maximum biomass yield, however, this bears multiple disadvantages. In particular, adaptive laboratory evolution favors strains that strive for the maximal growth rate instead of biomass yield. Furthermore, using a yield-based definition

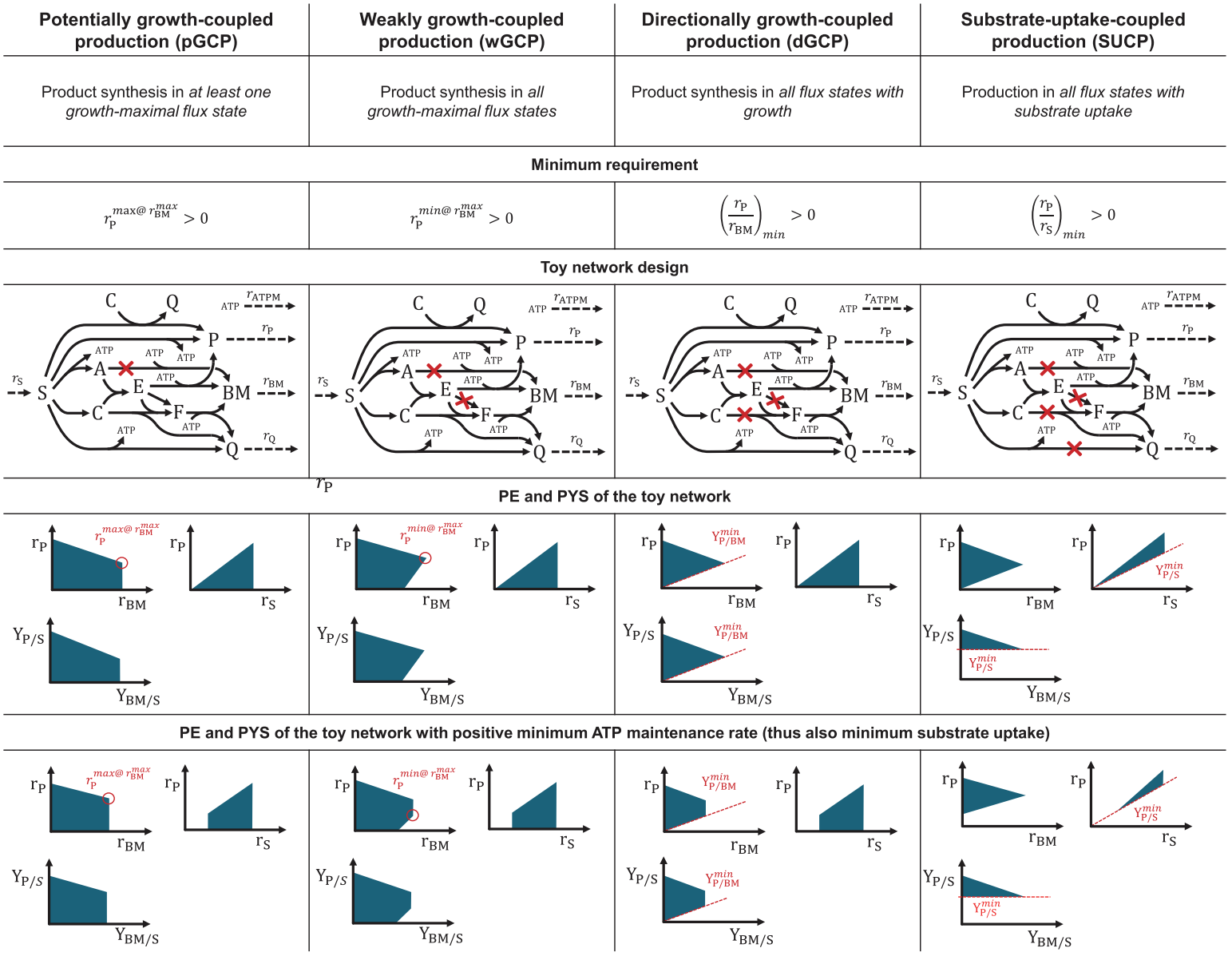

FIGURE 2 Overview of the four different degrees of growth-coupled product synthesis. The red crosses indicate suitable sets of knockouts that induce coupling between biomass (here represented by the biomass precursor BM) and product (P) synthesis with the respective coupling degree. Production envelopes and yield spaces are shown for the example network—with and without assumption of a minimum ATP maintenance demand constraint (implying a metabolic baseline activity). r_S: substrate uptake rate; r_P: product synthesis rate; r_{BM}: biomass production (growth) rate; r_{ATPM}: rate of ATP consumed for maintenance

may significantly complicate classical bilevel-based strain design approaches. A rate-based definition of wGCP is hence preferable.

There is a multitude of optimization methods developed for strain design for wGCP. These variants typically rely on bilevel optimization approaches, but differ in the optimization of certain aspects under weak coupling. For example, RobustKnock maximizes the lower bound of the product synthesis rate at the state of maximum growth,[3] which was alternatively achieved by using the OptKnock formulation with a tilted objective function.[2] More recent approaches, like OptORF variants[6] or OptCouple,[25] allow also the definition of a broader range of growth rates for which growth-coupling is enforced. In particular, OptCouple searches for interventions to maximize the distance between (a) the maximal growth rate with guaranteed production and (b) the maximal growth rate where product synthesis is not guaranteed.[25] However, despite their specific features, all these methods compute strain designs that share the minimal property of wGCP as defined in Equation (2).

As the third class of growth-coupling strength, we define directionally growth-coupled production (dGCP) to be present when growth implies product synthesis for any positive growth rate. Formally, this means

$$Y_{P/BM}^{min} = \left(\frac{r_P}{r_{BM}}\right)_{min} > 0, \tag{3}$$

that is, the "yield" of product per biomass, or, more precisely, the ratio of product and biomass synthesis rates, must be strictly positive. By definition, this ratio only exists (and is thus only relevant) for flux vectors with non-zero growth rate, and we therefore demand again that at least one flux vector with a non-zero growth rate exists. In the toy network, in addition to the knockouts of the pathways from A to BM and from E to F enforcing weak coupling, the pathway via F to BM also needs to be blocked. It has a reduced growth yield and might therefore not be relevant under optimal growth, but directional coupling demands that this pathway is also blocked as it allows growth without product synthesis. dGCP in a network can be ascertained by minimizing the ratio ("yield") of production and growth rate (such as yield minimization or maximization requires linear-fractional programming[33]). If the optimal value is greater than zero, dGCP is present. Likewise, dGCP can also be inferred graphically from the PE or PYS. In both representations, the entire flux space must be located above a diagonal line that may cross the x-axis only in the point of origin. The slope of this line is the minimum amount of product synthesized per amount of produced biomass. The term "directional coupling" as defined in flux coupling analysis[36]–[38] designates this coupling type and was therefore also used by Tervo and Reed (2014) in the context of strain design. In contrast, the terms full and partial coupling introduced in flux coupling analysis describe only special cases of the general directional coupling (see Figure 3). Potential and weak coupling as defined herein would be regarded as uncoupled in flux coupling analysis because the dependence between growth and product is only a local property.

We call the fourth and strongest type of coupling substrate-uptake-coupled production (SUCP), which demands that product synthesis

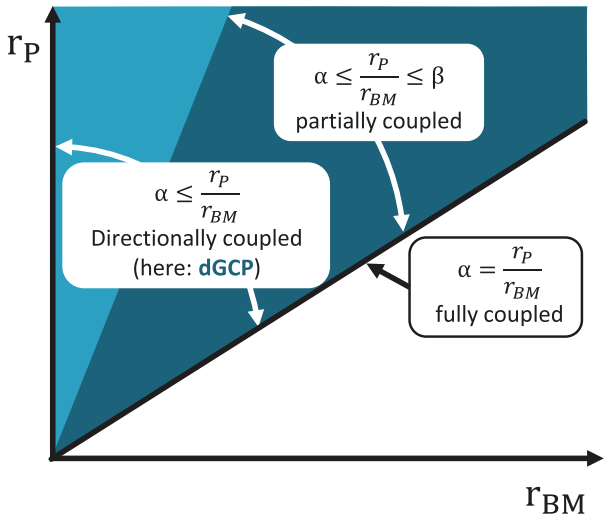

FIGURE 3 Different coupling types as defined in flux coupling analysis (cf. Burgard et al. 2004). Directional GCP requires a strictly positive ratio of r_P and r_{BM} in all flux vectors. $\alpha = \frac{r_P}{r_{BM}} > 0$ defines the minimum of this ratio and thus the slope of the outer black line

occurs in all flux states where substrate is taken up. This can be expressed as

$$Y_{P/S}^{min} = \left(\frac{r_P}{r_S}\right)_{min} > 0, \tag{4}$$

simply demanding that the yield of product per substrate is guaranteed to be above zero. In reality, this means that product synthesis occurs in all (non-zero) flux states, including those without growth. Even though the growth rate does not appear in condition (4), one may regard SUCP as GCP provided that growth is feasible: growth obviously requires substrate uptake, which in turn implies product synthesis due to (4). This is also the reason why SUCP has sometimes been called "strong growth coupling",[7,8,24,31,32] however, we would like to refrain from this term to make clear that the SUCP condition (4) as such does not require growth. In fact, condition (4) guarantees a minimum product yield even without growth and an SUCP strain design could also be used for two-stage processes with no or only little growth during the production phase (where growth is blocked, for example, by a genetic switch or nitrogen starvation; such a design approach has been called "non-growth production" in Garcia and Trinh (2019)[24]).

To achieve SUCP in the toy network, the route from substrate S to byproduct Q needs to be blocked in addition to the three knockouts of the dGCP as this pathway has a product (P) yield of zero (Figure 2). In the remaining network, all possible steady-state flux distributions will now definitely produce some P. Similar to directional growth-coupling, presence of SUCP can be confirmed through a product yield minimization. Accordingly, in the PYS, the entire flux space is located above a non-zero product yield threshold and does not touch the horizontal axis (corresponding to a product yield of zero) at any point. Importantly, as shown by the example in Figure 2, the PE under SUCP may have the same shape as under dGCP and is thus alone not conclusive whether SUCP is present. However, apart from the PYS, SUCP can also

TABLE 1 Mapping definitions of GCP used in previous studies to the four coupling degrees introduced in this work

Reference	Potentially growth-coupled production (pGCP)	Weakly growth-coupled production (wGCP)	Directionally growth-coupled production (dGCP)	Substrate-uptake-coupled production (SUCP)
[1]	OptKnock/Maximally coupled objectives	-	-	-
[42]	-	-	-	Full coupling/obligatory coupling
[3]	OptKnock	Obligatory byproduct of biomass formation	-	-
[2]	OptKnock/non-unique phenotype	Coupling	Full coupling	-
[6]	-	Weak coupling	Directional coupling	-
[16]	OptKnock/No (effective) growth-coupling	Partial coupling	Full coupling	Full coupling
[7]	-	Weak coupling	-	Strong coupling
[24,31]	-	Weak coupling (wGCP)	-	Strong coupling
[32]	OptKnock	Weak coupling	Holistic coupling	Strong coupling
[41]	-	Weak coupling (MCSw)	-	MCSf, MCSe

be identified in the two-dimensional projection where the axes represent substrate uptake rate (instead of growth rate used in the PE) and product synthesis rate. In this phase plane, in terms of flux coupling analysis, SUCP generally occurs as partial instead of directional coupling between r_P and r_S (Figure 2; cf. with Figure 3) because the latter would imply that substrate-independent production was possible. Furthermore, if a model carries a baseline metabolism that requires a minimum substrate uptake (e.g., due to some non-growth-associated ATP maintenance, often imposed in metabolic models), this, in turn, results in a minimum production rate (see Figure 2, last row).

As shown in the next section, the four criteria (1–4) can be formulated with tighter constraints demanding that the product synthesis rates in (1) and (2), or the product-per-biomass yield (3), or the product-per-substrate yield (4) must not fall below a certain non-zero threshold. This is indeed often used in the problem setups of strain design methods, likewise, constraints, for example, for a minimal possible growth rate, are often added. However, with the most relaxed versions (1–4), a clear hierarchical dependency exists between the different degrees of GCP: a network showing wGCP automatically fulfills the criterion for pGCP; if dGCP is present then the criteria of wGCP and pGCP are also naturally fulfilled, while a strain with SUCP satisfies the criteria of the other three types of GCP. In this regard, the actual GCP type of a concrete network (or strain) design is determined by the strongest of the four conditions (1–4), that is fulfilled in the network. For, example, a wGCP strain design fulfills the wGCP condition (2) but not the dGCP condition (3). Furthermore, it may happen that strain design solutions may exist for a weaker (e.g., wGCP) but not for a stronger (e.g., dGCP, SUCP) coupling type. In the opposite direction, this hierarchy also implies that methods computing strain designs for wGCP may, for instance, also return solutions that enforce dGCP or even SUCP. These hierarchical relationships will be further discussed in the Results section.

Since previous studies used different notions and terminology in the context of growth-coupled product synthesis, in Table 1, we sought to semantically map these previously used terms to the definitions introduced herein. Most of the previous works distinguished only two major types of GCP. As one exception, Alter and Ebert (2019) considered three different degrees of coupling (weak, holistic, strong) which are related with our definitions of wGCP, dGCP and SUCP, respectively. However, their definitions require the unnecessary assumption of a metabolic baseline flux and are purely based on the PE, which, as shown for dGCP and SUCP, can be ambiguous.

It should be noted that the four terms potentially/weakly/directionally GCP and substrate-uptake-coupled production describe properties that are either present or absent in a given (or designed) metabolic network model. However, the coupling terms themselves do not qualify as a benchmark for the actual production performance. For example, strains with directional GCP are not per se better production hosts than strains with weak GCP, nor vice versa. For assessing the suitability and production potential of designed strains, different measures can be used.[2,25,39-41]

2.2 | Using the MCS framework to compute strain designs for all four coupling degrees

The vast majority of constraint-based metabolic design algorithms use growth-coupled product synthesis as the underlying design principle. However, so far, there is no single computational framework that allows the computation of intervention strategies for all four types of coupling strengths in genome-scale networks. As already mentioned in the introduction section, bilevel (biased) optimization approaches naturally focus on (different variants of) pGCP and wGCP designs while MCS-related strain design calculations predominantly demanded

SUCP. Surprisingly, explicit computation of dGCP, as a medium coupling strength, has rarely been considered with the exception of two purely theoretical studies.[6,43] In the following, we will explain how the MCS framework can be used to compute metabolic designs for all four types of coupling. This will require an extension of the current MCS problem formulation to also allow computation of pGCP and wGCP designs in large-scale networks.

A detailed introduction to the theory of MCS can be found in literature[28,29,40,44-46] and in the Supplementary Text. For the calculation of MCS, it is fundamental to define a set of undesired (target) behaviors and (optionally) a set of desired (protected) behaviors. Both are specified through sets of linear inequalities. An MCS is then defined as a minimal (irreducible) set of interventions that block all undesired and preserve at least some desired behaviors. Interventions are typically reaction or gene knockouts, but reaction/gene additions or overexpression[47] can also be considered.[48] Initially, MCS were calculated from a given set of undesired and desired elementary modes. In fact, this approach allows the computation of metabolic designs for all four coupling types by selecting appropriate sets of desired and undesired elementary modes (e.g., wGCP and SUCP were computed in[29]). However, this approach becomes quickly infeasible in large-scale networks where elementary modes cannot be fully enumerated. Therefore, duality-based approaches have been developed allowing the computation of shortest MCS also in genome-scale networks in one step via mixed integer linear programming (MILP), without computation of elementary modes in a preprocessing step.[28] An introduction to the dual computation of MCS can be found in the Supplementary Text. Again, as a common principle for MCS-based methods, desired and undesired (target) behaviors must be specified, which, for the dual approach, is done in the form of linear inequalities. The computation of MCS for growth-coupled design requires the proper translation of the demands of a particular GCP type into linear inequality systems that describe the desired and undesired flux states.

For the case of SUCP, the undesired flux states are straightforward to describe. According to Equation (4), the MCS need to block all flux solutions that have a positive substrate uptake rate but a zero product synthesis rate, that is, a zero product yield:

$$\frac{r_P}{r_S} = 0, \tag{5}$$

The yield term (4) is actually non-linear and is only defined for strictly positive substrate uptake. To integrate this constraint in the MCS framework, it needs to be linearized via the two relations

$$r_P = 0 \tag{6}$$

$$r_S > 0. \tag{7}$$

Strict inequalities as in (7) cannot be used in linear programming, and so we replace (7) with:

$$r_S \geq \varepsilon, \tag{8}$$

with a sufficiently small number $\varepsilon > 0$ as a lower threshold for r_S. Due to numerical tolerances used by most solvers, too small numbers for ε should be avoided, otherwise the MCS algorithm will seek to also eliminate flux vectors of zero substrate uptake and no production. Importantly, if a metabolic network has some basal activity (e.g., due to a non-zero ATP maintenance demand), Equations (7) and (8) are automatically fulfilled and can be dropped, hence, only Equation (6) would remain.

As was already mentioned before, stronger constraints for the product yield are often demanded by specifying a minimum product yield threshold $Y_{P/S}^{Target} > 0$ for the right-hand side in Equation (4). In this case, Equation (5) would read

$$\frac{r_P}{r_S} \leq Y_{P/S}^{Target}, \tag{9}$$

which, under the constraint (8), can be safely rewritten to

$$r_P - Y_{P/S}^{Target} r_S \leq 0. \tag{10}$$

As desired (protected) region for SUCP, we demand that growth with a given threshold for a (desired) minimal growth rate should be possible:

$$r_{BM} \geq r_{BM}^{Desired}. \tag{11}$$

By demanding Equation (11) we do not need to explicitly demand the fulfillment of Equation (4) because the computed MCS will anyway ensure that all flux vectors with zero product yield (as defined by Equations (6) and (8)) will be blocked, hence, if there remains a feasible flux vector obeying (11), it will automatically satisfy Equation (4). To summarize, as also graphically illustrated in Figure 4, an MCS problem for enforcing SUCP is defined by the undesired flux vectors described by Equation (6) (or (10)) and (8) and the desired flux vectors specified by (11).

The procedure for directional GCP is analogous to SUCP. According to condition (3), for dGCP all flux states with growth but zero product synthesis must be blocked. Equations (5) and (7) now translate to:

$$r_P = 0 \tag{12}$$

$$r_{BM} \geq \varepsilon. \tag{13}$$

As a more general version, we can again replace (12) with a threshold for the "product-per-biomass yield":

$$r_P - Y_{P/BM}^{Target} r_{BM} \leq 0. \tag{14}$$

For the protected flux states, we can again use (11).

We may proceed in the same way to formulate the undesired space so as to obtain MCS for weak GCP using the coupling condition in Equation (2). Desired flux vectors can again be defined by (11). For the target flux vectors, analogous to SUCP and dGCP, Equation (2) can be

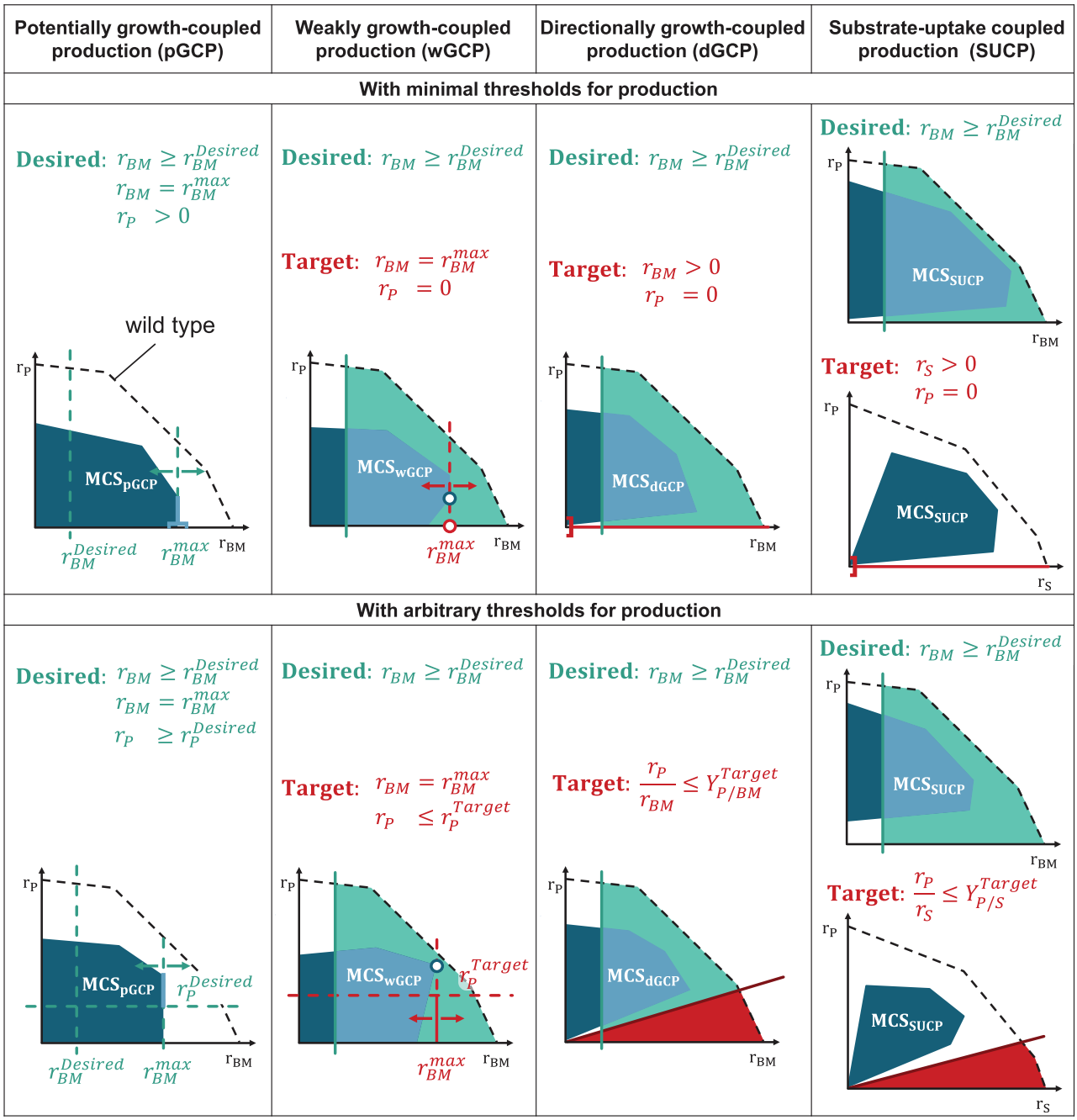

FIGURE 4 The MCS approach can be used to compute strain designs for all four coupling types by specifying suitable desired and undesired flux regions. For the computation of MCS strain designs inducing pGCP and wGCP, growth optimality in the desired/target regions is implicitly demanded in the optimization problems (see also Supplementary Text). Exemplary wild type flux spaces are outlined with a dashed black line, desired flux spaces are shown in green, target flux spaces in red, and the designed flux spaces (after implementing the MCS) in dark blue. The upper row shows the case with minimal coupling requirements and the lower row the most general case with arbitrary thresholds for $r_P^{Desired}$(pGCP), r_P^{Target}(wGCP), $Y_{BM/S}^{Target}$ (dGCP), and $Y_{P/S}^{Target}$ (SUCP). Note that all inequalities with a strict smaller sign (*e.g.*, $r_P > 0$) will be approximated by non-strict inequalities (e.g., $r_P \geq \varepsilon$) for the MCS computation

translated to the following two inequalities

$$r_P = 0 \tag{15}$$

$$r_{BM} = r_{BM}^{\max}, \tag{16}$$

to demand that all flux vectors need to be blocked that have no production at maximal growth. The problem here is that r_{BM}^{max} depends implicitly on the knockouts yet to be identified by the MCS algorithm and can thus not be defined a priori. In the Supplementary Text, we show how LP duality theory can be used to translate constraint

(16) into a (single-level) optimality constraint in the MCS computation framework.

The same approach finds application in the MCS setup for potentially GCP (see Supplementary Text). As a peculiarity, pGCP does not require the formulation of undesired behavior and only needs specification of a desired behavior. For this, we complement constraint (11) with the following two inequalities:

$$r_P > 0 \tag{17}$$

$$r_{BM} = r_{BM}^{max} \tag{18}$$

The optimality constraint (18) can again be resolved by duality theory and the strict inequality in (17) be approximated with a sufficiently small number $\varepsilon > 0$:

$$r_P \geq \varepsilon. \tag{19}$$

With the implicit integration of optimality constraints, all four coupling types can now be handled with the MCS framework as summarized in Figure 4 (for the simple conditions (1–4) as well as for the more general case with arbitrary thresholds for production rates and yields). In the Results section, we will make use of this scheme to compute and compare MCS for the four different coupling strengths in relevant application examples. The parameters used in each MCS computation are shown in Supplementary Table S1.

2.3 | Implementation, scripts, and models used for the example calculations

The models and scripts used in this work are provided at GitHub: https://github.com/klamt-lab/MCS_growth-coupling. The scripts require the recent version of *CellNetAnalyzer*[30,49] (2021.1), a freely available MATLAB toolbox. The new features for using MCS with implicit optimality constraints have been integrated in API functions. The genome-scale MCS computations were performed with MATLAB 2020b and IBM ILOG CPLEX 12.10 on single nodes of a high-performance cluster with two 8-core Intel Xeon Skylake Silver 4110 and 192 GB memory per node.

3 | RESULTS

We used the MCS framework to generate and compare exemplary knockout-based strain designs for all four defined coupling types (pGCP, wGCP, dGCP, and SUCP). With these calculations, we aim to demonstrate the nested hierarchy of strain designs with increasing coupling strength and to validate the newly developed MCS approach for computing wGCP and pGCP strain designs. We also assess whether using weaker coupling strengths (pGCP, wGCP or even dGCP) pays off in terms of number of required interventions and the computation time compared to the strongest variant SUCP. Finally, we show how the theory and variants of GCP can be generalized to other types of coupling strategies, for example, ATP-coupled product synthesis.

3.1 | Comparing strain designs for GCP of ethanol for all coupling strengths

In the first computation, we used an *E. coli* core model and enumerated gene MCS for GCP of ethanol for all four coupling degrees. The core model was derived from the genome-scale model *i*ML1515[50] analogous to *EColiCore2* derived from *iJO1366*[51] (reactions and species are listed in Table S2). The respective MCS problems were set up as described in the previous section using the most relaxed formulations (upper row in Figure 4). In order to compute comparable sets of MCS, we used identical thresholds for the demanded minimum growth rate ($r_{BM}^{Desired} \geq 0.05\ h^{-1}$) in the desired regions. In this way, MCS for the four coupling types are expected to exhibit the aforementioned nested hierarchy. In addition to gene-knockouts, the MCS were allowed to block oxygen supply as an intervention. We considered two scenarios (cf. also Figure 2): scenario (A), where a minimum ATP demand ($r_{ATPM} \geq 6.86\ mmol\ g_{CDW}^{-1} h^{-1}$) for non-growth associated maintenance (NGAM) processes was included (which is standard in most models), and scenario (B) without a minimum NGAM demand for ATP ($r_{ATPM} \geq 0$).

For simplicity, Figure 5 shows the results for the MCS up to size 3 for both scenarios. The chosen representation highlights the hierarchical relationships between the MCS for the different coupling types. This hierarchy implies, for example, that an MCS computed for pGCP might simultaneously fulfill the requirements for the stronger coupling types wGCP, dGCP, or SUCP. Likewise, an MCS for wGCP may also be a valid MCS enforcing dGCP and SUCP and an MCS computed for dGCP may also ensure SUCP. On the other hand, if, for example, an MCS calculated for wGCP does not imply dGCP (or SUCP) then it might be extendable, with additional knockouts to obtain a valid dGCP MCS. Yet, in the reverse direction, every MCS for dGCP can be inherited from at least one wGCP MCS, that is, it is either identical to or a superset (with additional knockouts) of an MCS found for wGCP. The same hierarchical relationship holds for any pair of coupling degrees. For example, wGCP and pGCP for ethanol synthesis can be established in both scenarios by blocking only the oxygen supply, while for dGCP and SUCP additional knockouts are needed. In fact, with at least one additional knockout (e.g., ldhA, the lactate dehydrogenase gene), it is possible to extend the pGCP- and wGCP-related MCS (O_2) to an dGCP MCS (O_2, ldhA) (note that net CO_2 uptake was not allowed in the model, otherwise fumarate reduction needs also to be blocked). In turn, such an dGCP MCS might also be a valid SUCP MCS (as is the case for (O_2, ldhA)) and if not, then it might be extendable to an SUCP MCS. For example, in the scenario with zero ATP maintenance demand (Figure 4B), the dGCP MCS (O_2, atpS) is not sufficient for SUCP, but it can be extended with a knockout of the glucose PTS (via the genes of ptsI or ptsH) to reach SUCP. Interestingly, these two SUCP MCS do not exist in scenario (A) because here knocking out the PTS would reduce the maximal growth rate below the given minimum threshold due to the posed ATP NGAM demand in scenario (A). For the same reason, there are several dGCP MCS with

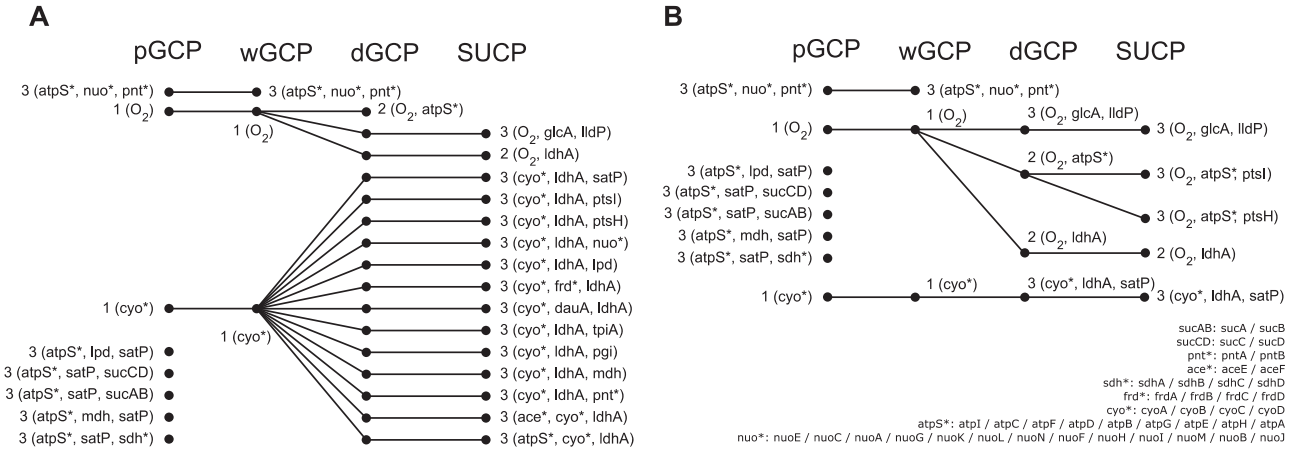

FIGURE 5 Summary of all gene MCS up to size 3 for growth-coupled production of ethanol in a metabolic core model of *E. coli* generated from two different scenarios. (A) with and (B) without a minimum NGAM ATP demand. Each black dot represents an MCS, the numbers indicate the size of the MCS and the lines the hierarchical relationships between the MCS of the different coupling degrees. Tables S3 and S4 list all MCS and their characteristics

two knockouts in scenario (B) which are not applicable for scenario (A). This demonstrates that the specified minimum ATP NGAM demand in metabolic models has a significant effect on the MCS and thus needs careful consideration.

Except for the single case described above, we found that all other dGCP MCS are also valid SUCP MCS in both scenarios. In contrast, not a single pGCP or wGCP MCS was a valid dGCP or even SUCP MCS. Moreover, there is one MCS valid for pGCP and wGCP (atpS, nuo, pnt), which is not extendable to an MCS for dGCP, thus occurring as "dead end" in Figures 5A and 5B. Likewise, a number of MCS with three knockouts exist for pGCP that do not reappear, neither identically nor as supersets, for stricter coupling types. There are three possible reasons why supersets of such MCS may not be found as solution for stricter coupling. First, supersets may exist, but with a higher cardinality than allowed in the MCS enumeration. Second, knockouts, which would increase the coupling strength, might be prohibited in the computation, as is the case for spontaneous reactions. Lastly, an MCS for stronger coupling may simply not exist due to the network topology and constraints, for example, when a reaction with a strictly positive lower bounded rate, like ATP maintenance, cannot be coupled with production.

3.2 | Genome-scale calculations for selected products

Next, we applied the extended MCS framework to compute realistic growth-coupled strain designs in a genome-scale model of *E. coli* (*i*ML1515[50]) for several relevant target products. The goal was to compare the number of required interventions and the computational effort to determine strain designs for the four different coupling degrees. We considered the production of 12 different (native and heterologous) products. For heterologous products, the required pathways were added to the model (detailed pathways listed in the Table S5). In contrast to the calculations in the previous chapter, where the strain design just had to exclude zero production, we here demanded higher minimum production thresholds, since predominantly poorly performing strain designs are typically found when using low production minima.[25,52] We used a reference production rate $r_{P,ref}$ to define the respective target region for each coupling degree. $r_{P,ref}$ was determined as 20% of the maximum possible production when the wild type strain, extended with heterologous pathways if applicable, grows at 20% of its maximal growth rate. For pGCP we demanded that this reference production rate ($r_{P,ref}$) would be attainable together with the maximum growth rate of the MCS strain. For wGCP we defined the undesired fluxes as those with production inferior to the threshold $r_P \leq r_{P,ref}$ at maximum growth rate. For dGCP we demanded a minimum ratio of growth and production rate by targeting all fluxes below the threshold $\frac{r_P}{r_{BM}} \leq \frac{r_{P,ref}}{0.2\, r_{BM,max\ (WT)}}$. Finally, in the case of SUCP we targeted fluxes with $\frac{r_P}{r_S} \leq \frac{r_{P,ref}}{r_{S,max}}$. In all setups, we additionally specified a desired region to guarantee that found strain designs still allow for growth rates greater than 0.05 h^{-1}. For the NGAM demand of ATP we used $r_{ATPM} \geq 6.86\ mmol\ g_{CDW}^{-1} h^{-1}$. The parameters for each computation are listed in Table S1.

In contrast to the full enumeration used in the *E. coli* core model above, we applied the following two procedures for finding single MCS, which are also available in *CellNetAnalyzer*: In the first computation we aimed to find (quickly) just a single MCS, without the necessity to obtain the MCS with the smallest number of knockouts (Table 2), while in the second procedure we searched for an MCS with a low (ideally minimum number) of knockouts (Table 3). We emphasize that the performance of these calculations depends on many factors, such as choice of the MILP solver (here CPLEX) and its parameters. Hence, the presented results of the computations provide a small sample that only exposes tendencies. To at least reduce dependency of the results on the chosen seed, we performed 12 runs in the first scenario

TABLE 2 MCS strain designs for the growth-coupled (pGCP, wGCP, dGCP, SUCP) production of 12 different products with arbitrary number of knockouts

	pGCP			wGCP			dGCP			SUCP		
Product	**av. size**	**suc/tot**	**av. runt.**	**av. size**	**suc/tot**	**av. runt.**	**av. size**	**suc/tot**	**av. runt.**	**av. size**	**suc/tot**	**av. runt.**
Ethanol	1.0	12/12	4 min	1.0	12/12	4 min	2.6	11/12	4 min	4.3	12/12	4 min
Lysine	13.0	2/12	10 min	-	0/12	timeout	13.0	1/12	9 min	25.1	8/12	14 min
Glutamate	4.5	12/12	4 min	7.0	4/12	5 min	12.0	1/12	7 min	17.3	8/12	6 min
Isobutanol (h)	7.2	10/12	5 min	4.9	7/12	8 min	7.3	6/12	13 min	10.1	12/12	4 min
1,4-Butanediol (h)	5.4	9/12	18 min	4.5	4/12	4 min	10.8	5/12	4 min	8.3	12/12	5 min
2,3-Butanediol (h)	6.3	12/12	4 min	7.2	10/12	8 min	8.9	7/12	19 min	8.4	12/12	4 min
Itaconic acid (h)	5.6	12/12	4 min	9.0	3/12	9 min	9.9	8/12	10 min	12.8	12/12	4 min
Isoprene (h)	12.1	9/12	14 min	11.3	3/12	41 min	10.7	3/12	13 min	19.9	8/12	11 min
Butane (h)	6.5	10/12	11 min	6.3	4/12	7 min	11.0	4/12	5 min	24.0	3/12	14 min
Methacrylic acid (h)	8.2	12/12	4 min	12.0	5/12	7 min	10.3	3/12	5 min	10.6	12/12	6 min
Resveratrol (h)	8.5	11/12	10 min	9.0	2/12	60 min	14.2	5/12	22 min	18.8	6/12	8 min
Bisabolene (h)	10.9	8/12	20 min	10.9	7/12	26 min	11.8	5/12	4 min	12.6	11/12	10 min
Mean (all rows except lysine)	**6.9**	**10.6/12**	**9 min**	**7.6**	**5.5/12**	**16 min**	**10.5**	**5.3/12**	**10 min**	**13.4**	**9.8/12**	**7 min**

In these calculations, we aimed to find just a single MCS, without the necessity to obtain the smallest one. For each combination of product and growth-coupling degree, 12 computations were started with different seeds and with a time limit of 2 h each (details for each run are shown in Table S6). The last row contains the mean values over all scenarios except for lysine where a meaningful comparison is not possible because no MCS could be found for wGCP. Computed MCS are shown in Table S7.

Abbreviations: av. size, average number of knockouts per MCS; suc/tot, number of successful computations (that returned an MCS) per total runs; av. runt, average runtime of successful computations; (h), next to product name marks heterologous products

with a runtime limit of 2 h and six runs in the second procedure with a limit of 4 h (all with identical solver parameters but starting the computation from different seeds) for each combination of product and growth-coupling degree to obtain averaged values for runtime and MCS size (Tables 2 and 3). The code that was used for the computations is available on GitHub (see Implementation section).

We first analyze the data for the procedure that quickly searches for any MCS (without the necessity to find one with the smallest cardinality; (Table 2)). With the exception of lysine under wGCP, at least one MCS for all combinations of products and coupling degrees could be found. However, a more differentiated view can be obtained by looking at the number of successful runs (finishing before the timeout of 2 h) among the 12 different seeds for each product. For every coupling degree, we can find products, where at least one run did not succeed. Furthermore, it can be seen that the mildest (pGCP) and the strongest (SUCP) coupling degree have the highest success rates. For the runs that were successful we can see that, in average, SUCP was the fastest followed by pGCP and dGCP (similar) and with wGCP being the slowest. Finally, the computations also show a hierarchy in the sense that the average size of the MCS increases with coupling strengths, however, it should be noted that this is not necessarily the case since minimum cardinality for the MCS computations was here not demanded and there are indeed cases where the MCS of a stronger coupling degree have lower average size.

When searching for MCS with low (ideally minimum) cardinality shown in Table 3, we can see that only in relatively few cases MCS with guaranteed minimum number of knockouts could be found. A minimum solution could be found for all coupling degrees for the products ethanol, isobutanol, 1,4-butanediol and 2,3-butanediol. pGCP and SUCP show again a slightly better performance regarding successful finishing full minimization. On the other hand, dGCP and SUCP calculations were the only ones that could find for each product at least one MCS with (relatively) low number of interventions (e.g., pGCP and wGCP did not find solutions for lysine, wGCP furthermore missed resveratrol). However, pGCP was the fastest in cases where it finished the computation. wGCP and dGCP again had a higher number of timeouts (unsuccessful runs). In cases where the minimum MCS was found the monotone increase in MCS size with stronger coupling degrees can again be observed (but notice the larger MCS for isoprene for pGCP and wGCP (6) compared to dGCP (5) which occurs because the full minimization could not be finished in those cases).

3.3 | Generalization of coupled product synthesis: From growth-coupled to ATP-coupled production

So far, we focused on GCP of a target metabolite, the most common principle for strain design. In the following, we will show that this coupling principle can be generalized to other biological functions than growth. One relevant example is to couple ATP synthesis with product formation. Such a coupling might be particularly relevant for the idea of enforced ATP wasting as a metabolic engineering strategy, which

TABLE 3 MCS strain designs for the growth-coupled (pGCP, wGCP, dGCP, SUCP) production of 12 different products with minimum number of knockouts

	pGCP			wGCP			dGCP			SUCP		
Product	m. size	opt/suc/tot	av. runt.	m. size	opt/suc/tot	av. runt.	m. size	opt/suc/tot	av. runt.	m. size	opt/suc/tot	av. runt.
Ethanol	1	6/6/6	4 min	1	6/6/6	4 min	2	6/6/6	5 min	3	6/6/6	5 min
Lysine	-	0/0/6	timeout	-	0/0/6	timeout	11	0/4/6	timeout	11	0/6/6	timeout
Glutamate	3	6/6/6	38 min	7	0/1/6	timeout	8	0/2/6	timeout	8	0/6/6	timeout
Isobutanol (h)	2	6/6/6	6 min	4	1/2/6	3 h	4	2/2/6	3 h	5	2/6/6	4 h
1,4-Butanediol (h)	2	6/6/6	7 min	2	6/6/6	7 min	4	6/6/6	2 h	5	3/6/6	3 h
2,3-Butanediol (h)	3	6/6/6	31 min	4	1/2/6	86 min	4	2/6/6	77 min	4	6/6/6	30 min
Itaconic acid (h)	3	6/6/6	14 min	7	0/1/6	timeout	8	0/3/6	timeout	8	0/6/6	timeout
Isoprene (h)	6	0/4/6	timeout	6	0/2/6	timeout	5	0/5/6	timeout	7	0/6/6	timeout
Butane (h)	5	0/5/6	timeout	5	0/1/6	timeout	5	0/6/6	timeout	5	2/6/6	4 h
Methacrylic acid (h)	3	6/6/6	33 min	6	0/2/6	timeout	6	0/5/6	timeout	6	0/6/6	timeout
Resveratrol (h)	-	0/0/6	timeout	7	0/2/6	timeout	8	0/4/6	timeout	8	0/6/6	timeout
Bisabolene (h)	6	0/4/6	timeout	5	0/5/6	timeout	6	0/6/6	timeout	7	0/6/6	timeout
Mean (all rows except lysine and resveratrol)	**3.4**	**4.2/5.5/6**		**4.7**	**1.4/2.8/6**		**5.2**	**1.6/4.7/6**		**5.8**	**1.9/6/6**	

In these calculations, we aimed to find a single MCS with a low number of knockouts (ideally with the minimum number of knockouts). For each combination of product and growth-coupling strength, six computations were started with different seeds and with a time limit of 4 h each (details for each run are shown in Table S8). The last row contains the mean values of all scenarios except lysine and resveratrol, for which pGCP (and wGCP) MCS could not be found. Due to the large number of timeouts, the runtimes are not considered. Computed MCS are shown in Table S9.

Abbreviations: m. size, minimum size of all found MCS (smallest MCS); opt/suc/tot, number of computations that returned the smallest MCS/number of computations that returned an MCS without guaranteeing for minimality/total runs; av. runt, average runtime of computations with assured smallest MCS; (h), next to product name marks heterologous products.

received increased attention in recent literature.[53–60] The idea is that introduction of an ATP wasting mechanism, such as artificial futile cycles[53,55,57] or, more directly, via the ATP-hydrolyzing F_1-portion of the ATPase,[59-61] may boost substrate uptake and product synthesis if ATP production is coupled with the synthesis of the target metabolite. This could be particularly relevant for improving the performance of two-stage processes.[43]

It is now straightforward to consider the same four coupling strengths as for GCP. Potentially ATP-coupled production (pACP) would demand that product synthesis is possible under maximal ATP production. Weakly ATP-coupled production (wACP) means that product synthesis is mandatory under maximal ATP production. Directionally ATP-coupled production requires product synthesis whenever ATP is synthesized and substrate-uptake-coupled production again implies product formation whenever substrate is taken up. For meaningful definitions, we would demand in all four cases that ATP production is feasible in the network (similar to growth in the growth-coupled case). ATP-coupled strain designs can then be calculated in an analogous way to growth-coupled strain designs, for example, via MCS, by replacing terms with the growth rate with terms for ATP production. Net ATP production (and consumption) can be simulated via the NGAM reaction (r_{ATPM}) included in most models.

While this generalization can directly be applied, as it could be for coupling any other flux with product formation, the case of ATP synthesis requires special attention. While the consideration of pGCP and wGCP is meaningful in the context of (laboratory) evolution, it is less reasonable to assume that the cell strives to maximize ATP synthesis for non-growth-associated processes. For this reason, we will not further consider pACP and wACP. Second, under the premise that feasibility of growth and net ATP synthesis is ensured, the definition of SUCP is identical for growth- and ATP coupling as it implies product synthesis when substrate is taken up, which is a requirement for growth and ATP synthesis. This has some important consequences. In particular, the existence of SUCP-like growth-coupled strain designs proven for a wide range of products and production hosts[8] now also implies wide feasibility of ATP-coupled strain designs under SUCP because a minimum ATP maintenance constraint for the r_{ATPM} reaction was used in these calculations.

For the reasons given above we need not consider pACP, wACP, and SUCP for ATP-coupled product synthesis and focus now on directional ACP. Generally, since ATP is required for growth, dACP implies almost always dGCP. Only in "pathological" examples, growth may abolish ATP-coupled product synthesis. This can only happen if growth consumes ATP and the product (or a product precursor) in the same fixed ratio as both are generated. Such a case is extremely unlikely. We did not observe this in our computation examples, and we therefore assume in the following that dGCP is present in a strain whenever dACP is. However, the other way around does not necessarily hold.

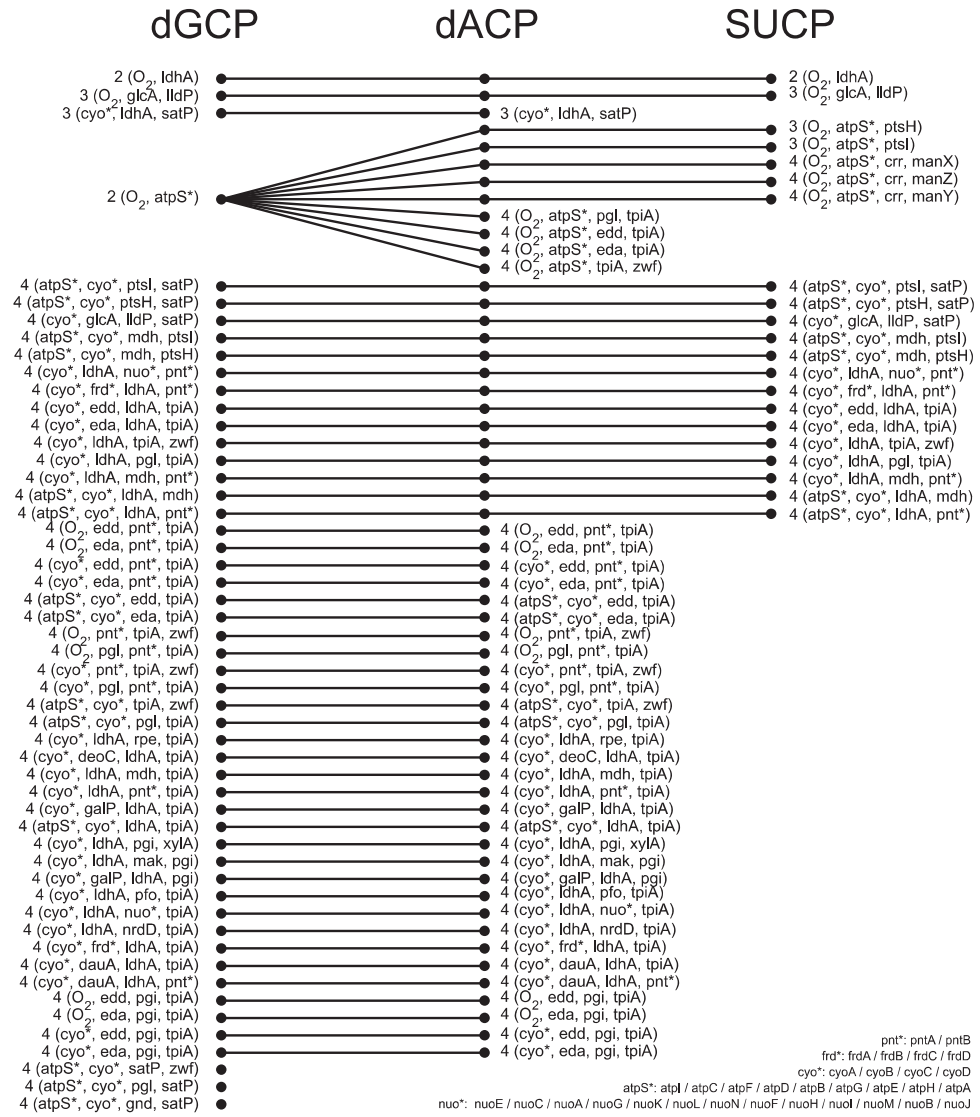

FIGURE 6 Comparison of MCS for dGCP, for dACP, and for SUCP for ethanol synthesis in the *i*ML1515 core model. In all cases, no NGAM was demanded. The MCS of size 2 and 3 for dGCP are identical with those from Figure 5B. ATP-coupled production takes an intermediate position between dGCP and SUCP. All SUCP-MCS also have dACP, which in turn have dGCP. Some MCS for dGCP are not ATP-coupled. In these cases, production is forced through alternative mechanisms

For example, an dGCP strain might be established because a target metabolite becomes, by suitable interventions, a necessary byproduct of a biosynthetic pathway and must then be excreted. This strain would not show dACP. Here we can again distinguish the two cases with and without a given demand of ATP for NGAM. Importantly, if the NGAM demand is not zero, all resulting strain designs computed for dACP will also fulfill the condition of SUCP: then, a minimum amount of substrate must be taken up to produce the demanded amount of ATP and since dACP demands some product synthesis whenever ATP is synthesized, dACP implies product synthesis whenever substrate is taken up. We can therefore concentrate on the case for dACP without a minimum NGAM demand.

For a comparison with the respective dGCP and SUCP strains, we fully enumerated MCS up to the size of 4 in the core model of *i*ML1515 for dACP of ethanol by *E. coli*. In analogy to the Equations (12) and (13) used for dGCP, the target system for dACP reads

$$r_P = 0 \tag{20}$$

$$r_{ATPM} \geq \varepsilon \tag{21}$$

and we also demanded feasibility of growth (Equation (11)).

The results of these computations are shown in Figure 6 and indicate that dACP takes an intermediate role within the (hierarchical) tree of MCS connecting dGCP with SUCP strain designs: all MCS for SUCP imply dACP which in turn implies dGCP. With the explanations given above, this was to be expected, first, because SUCP with feasible growth and net ATP formation implies both dACP and dGCP and, second, because directional coupling of ATP synthesis with ethanol

formation should also (directionally) couple growth with ethanol production since ATP is required for growth. In the reverse direction, Figure 6 suggests that many dGCP strategies are induced by directional coupling of ATP synthesis with ethanol formation, and that some dACP strategies in turn even fulfill SUCP. An example of an MCS that is valid for all strategies is to block supply of oxygen and to knockout the pathway to the alternative fermentation product lactate. However, There are some MCS for dGCP that are not valid dACP strategies (e.g., fourth MCS and last MCS in Figure 6) indicating that the mechanism of growth-coupling induced by these MCS must be ATP-independent. For example, as suggested by other studies, branching points in the metabolism, known as anchor reactions, may serve as suitable targets to make a compound a mandatory by-product of essential biomass precursors.[32,62] The fourth dGCP MCS in Figure 6 is a particularly interesting case as this MCS can be extended to (directionally) couple both biomass and ATP synthesis with product formation and that some (but not all) of these dACP strategies directly support SUCP.

4 | DISCUSSION

Coupling growth with product synthesis is the most common design principle used in computational strain design for metabolic engineering. In this study, we reviewed and systematized the existing notions resulting in four unambiguous classes. All existing methods for computing growth-coupled strategies can be assigned to one of these four categories. We based each coupling type on simple mathematical definitions implying a clear hierarchical order, such that a strain design strategy for a stronger coupling type automatically implies all weaker ones. Simple constraint-based modeling techniques such as FBA, PEs, and yield spaces can be used to test whether coupling of a certain type is present in a given network (design).

We clarified how the framework of MCS can be employed to compute strain design for each of the four classes. While the formulation of specific MCS problems to find dGCP and SUCP strain designs is straightforward (and has, for the case of SUCP, been used in previous work) strategies for pGCP and wGCP could so far not be computed with MCS. We therefore extended the existing MCS framework with implicit constraints for the maximal growth rate, so that product synthesis can now be demanded for the special case of optimal growth. This extension closes the gap between MCS-based and bilevel-based strain design approaches. The new optimality constraint feature for (reaction or gene) MCS was also implemented in *CellNetAnalyzer* and is compatible with most other recent MCS developments, such as the definition of multiple target and desired regions as well as substrate and heterologous pathway additions.[48]

We illustrated the hierarchical dependency of the four coupling types by exemplarily computing and analyzing MCS for GCP of ethanol in *E. coli*. The results confirmed that an MCS found for a demanded stronger coupling degree is always a superset of (or identical to) at least one MCS computed for weaker coupling. In the reverse direction, searching for MCS of a weaker coupling type (e.g., dGCP) may deliver strain design strategies that are also valid for stronger coupling (e.g., SUCP). If a valid MCS found for a given coupling type (e.g., dGCP) does not imply even stronger coupling (e.g., SUCP), then it might be extensible to an MCS of stronger coupling by adding further interventions. In a second computation, we used MCS to compute realistic gene-knockout strategies that prime *E. coli* for the production for 12 native and heterologous products, again using all four coupling types. We found that in most cases where MCS with a minimum number of cuts could be found, demanding weaker coupling strengths may be beneficial in terms of required interventions, especially when comparing pGCP against the other three coupling strengths. However, much less is saved (typically only one or two interventions) when comparing the smallest number of interventions in wGCP and dGCP against SUCP (see last row in Table 3). Moreover, although the respective runs always depend on the problem setup and the chosen solver (parameters) and seeds, we noted a tendency that computing the respective MCS for pGCP and SUCP had higher success rates compared to wGCP and dGCP and was also faster (in case of SUCP at least when searching for any MCS; Table 2). For wGCP, this can be explained by the fact that the strain design optimization problem has the largest size of all coupling degrees (see Supplementary Text). Also, we did not encounter any case where a solution for the weakest coupling degree (pGCP) existed but not for the strongest (SUCP). This confirms results of a recent study showing that SUCP strain designs exist for almost all potential products (metabolites) in five important production hosts.[8] Altogether, from the perspective of computability, pGCP and SUCP seem to be preferable, however, since pGCP cannot guarantee product synthesis (even not at optimal growth) this could be an argument for favoring SUCP strategies. Generally, we recommend assessing the feasibility of each coupling type by computing (single) smallest or even random MCS before searching for an MCS with minimum number of interventions.

As a last methodological development, we showed how product synthesis can be coupled to other biological functions than growth and that the notion of coupling strengths as introduced herein for growth coupling can be naturally generalized for those cases. As a relevant example for metabolic engineering, especially in the context of enforced ATP wasting strategies, we discussed the design principle of ACP. Since ATP synthesis is directly relevant for growth, there are several relationships between growth-coupled and ATP-coupled strain designs. In particular, intervention strategies inducing strongly ATP-coupled product synthesis imply practically in all cases also dGCP while the converse is not necessarily true.

ACKNOWLEDGMENTS

This work was funded by the European Research Council (721176) and MITACS.

Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERETS

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

P. S.: Conceptualization; formal analysis; investigation; methodology; software; visualization; writing-original draft; writing-review & editing. R. M.: Conceptualization; formal analysis; funding acquisition; methodology; project administration; resources; supervision; writing-review & editing. S. K.: Conceptualization; formal analysis; funding acquisition; investigation; methodology; project administration; supervision; writing-original draft; writing-review & editing

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article. The models and scripts used in this work are provided at GitHub: https://github.com/klamt-lab/MCS_growth-coupling.

ORCID

Philipp Schneider https://orcid.org/0000-0001-8858-428X
Radhakrishnan Mahadevan https://orcid.org/0000-0002-1270-9063
Steffen Klamt https://orcid.org/0000-0003-2563-7561

REFERENCES

1. Burgard, A. P., Pharkya, P., & Maranas, C. D. (2003). Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, *84*(6), 647–657. https://doi.org/10.1002/bit.10803
2. Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., & Palsson, B. Ø. (2010). Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metabolic Engineering*, *12*(3), 173–186. https://doi.org/10.1016/j.ymben.2009.10.003
3. Tepper, N., & Shlomi, T. (2010). Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, *26*(4), 536–543. https://doi.org/10.1093/bioinformatics/btp704
4. Otero, J. M., Cimini, D., Patil, K. R., Poulsen, S. G., Olsson, L., & Nielsen, J. (2013). Industrial Systems Biology of Saccharomyces cerevisiae Enables Novel Succinic Acid Cell Factory. *Plos One*, *8*(1), e54144. https://doi.org/10.1371/journal.pone.0054144
5. Campodonico, M. A., Andrews, B. A., Asenjo, J. A., Palsson, B. O., & Feist, A. M. (2014). Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metabolic Engineering*, *25*, 140–158. https://doi.org/10.1016/j.ymben.2014.07.009
6. Tervo, C. J., & Reed, J. L. (2014). Expanding metabolic engineering algorithms using feasible space and shadow price constraint modules. *Metabolic Engineering Communications*, *1*, 1–11. https://doi.org/10.1016/j.meteno.2014.06.001
7. Klamt, S., & Mahadevan, R. (2015). On the feasibility of growth-coupled product synthesis in microbial strains. *Metabolic Engineering*, *30*, 166–178. https://doi.org/10.1016/j.ymben.2015.05.006
8. von Kamp, A., & Klamt, S. (2017). Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nature Communications*, *8*, 15956. https://doi.org/10.1038/ncomms15956
9. Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., & Palsson, B. O. (2005). In silico design and adaptive evolution of *Escherichia coli* for production of lactic acid. *Biotechnology and Bioengineering*, *91*(5), 643–648. https://doi.org/10.1002/bit.20542
10. Trinh, C. T., & Srienc, F. (2009). Metabolic engineering of Escherichia coli for efficient conversion of glycerol to ethanol. *Applied and Environmental Microbiology*, *75*(21), 6696–6705. https://doi.org/10.1128/AEM.00670-09
11. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., & Van Dien, S. (2011). Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology*, *7*(7), 445–452. https://doi.org/10.1038/nchembio.580
12. Ng, C., Jung, M., Lee, J., & Oh, M.-K. (2012). Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering. *Microbial Cell Factories*, *11*(1), 68. https://doi.org/10.1186/1475-2859-11-68
13. Layton, D. S., & Trinh, C. T. (2014). Engineering modular ester fermentative pathways in Escherichia coli. *Metabolic Engineering*, *26*, 77–88. https://doi.org/10.1016/j.ymben.2014.09.006
14. Harder, B.-J., Bettenbrock, K., & Klamt, S. (2016). Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metabolic Engineering*, *38*, 29–37. https://doi.org/10.1016/j.ymben.2016.05.008
15. Banerjee, D., Eng, T., Lau, A. K., Sasaki, Y., Wang, B., Chen, Y., Prahl, J.-P., Singan, V. R., Herbert, R. A., Liu, Y., Tanjore, D., Petzold, C. J., Keasling, J. D., & Mukhopadhyay, A. (2020). Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. *Nature Communications*, *11*(1), 5385. https://doi.org/10.1038/s41467-020-19171-4
16. Machado, D., & Herrgård, M. J. (2015). Co-evolution of strain design methods based on flux balance and elementary mode analysis. *Metabolic Engineering Communications*, *2*, 85–92. https://doi.org/10.1016/j.meteno.2015.04.001
17. Maia, P., Rocha, M., & Rocha, I. (2016). In silico constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiology and Molecular Biology Reviews*, *80*(1), 45–67. https://doi.org/10.1128/MMBR.00014-15
18. Sandberg, T. E., Salazar, M. J., Weng, L. L., Palsson, B. O., & Feist, A. M. (2019). The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metabolic Engineering*, *56*, 1–16. https://doi.org/10.1016/j.ymben.2019.08.004
19. Maranas, C. D., & Zomorrodi, A. R. (2016). Computational Strain Design. In Maranas, C. D. & A. R. Zomorrodi, *Optimization Methods in Metabolic Networks* (pp. 155–172). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119188902.ch8
20. Pharkya, P., Burgard, A. P., & Maranas, C. D. (2004). OptStrain: A computational framework for redesign of microbial production systems. *Genome Research*, *14*(11), 2367–2376. https://doi.org/10.1101/gr.2872004
21. Kim, J., Reed, J. L., & Maravelias, C. T. (2011). Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *Plos One*, *6*(9), e24162. https://doi.org/10.1371/journal.pone.0024162
22. Tervo, C. J., & Reed, J. L. (2012). FOCAL: An experimental design tool for systematizing metabolic discoveries and model development. *Genome Biology*, *13*(12), R116. https://doi.org/10.1186/gb-2012-13-12-r116
23. Alter, T. B., Blank, L. M., & Ebert, B. E. (2018). Genetic optimization algorithm for metabolic engineering revisited. *Metabolites*, *8*(2), 33. https://doi.org/10.3390/metabo8020033
24. Garcia, S., & Trinh, C. T. (2019). Multiobjective strain design: A framework for modular cell engineering. *Metabolic Engineering*, *51*, 110–120. https://doi.org/10.1016/j.ymben.2018.09.003
25. Jensen, K., Broeken, V., Hansen, A. S. L., Sonnenschein, N., & Herrgård, M. J. (2019). OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled

strain designs. *Metabolic Engineering Communications*, *8*, e00087. https://doi.org/10.1016/j.mec.2019.e00087
26. Trinh, C. T., Unrean, P., & Srienc, F. (2008). Minimal Escherichia coli cell for the most efficient production of ethanol from hexoses and pentoses. *Applied and Environmental Microbiology*, *74*(12), 3634–3643. https://doi.org/10.1128/AEM.02708-07
27. Trinh, C. T., Liu, Y., & Conner, D. J. (2015). Rational design of efficient modular cells. *Metabolic Engineering*, *32*, 220–231. https://doi.org/10.1016/j.ymben.2015.10.005
28. von Kamp, A., & Klamt, S. (2014). Enumeration of smallest intervention strategies in genome-scale metabolic networks. *Plos Computational Biology*, *10*(1), e1003378. https://doi.org/10.1371/journal.pcbi.1003378
29. Hädicke, O., & Klamt, S. (2011). Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metabolic Engineering*, *13*(2), 204–213. https://doi.org/10.1016/j.ymben.2010.12.004
30. von Kamp, A., Thiele, S., Hädicke, O., & Klamt, S. (2017). Use of CellNetAnalyzer in biotechnology and metabolic engineering. *Journal of Biotechnology*, *261*, 221–228. https://doi.org/10.1016/j.jbiotec.2017.05.001
31. Garcia, S., & Trinh, C. T. (2020). Harnessing natural modularity of metabolism with goal attainment optimization to design a modular chassis cell for production of diverse chemicals. *ACS Synthetic Biology*, *9*(7), 1665–1681. https://doi.org/10.1021/acssynbio.9b00518
32. Alter, T. B., & Ebert, B. E. (2019). Determination of growth-coupling strategies and their underlying principles. *Bmc Bioinformatics*, *20*(1), 447. https://doi.org/10.1186/s12859-019-2946-7
33. Klamt, S., Müller, S., Regensburger, G., & Zanghellini, J. (2018). A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metabolic Engineering*, *47*, 153–169. https://doi.org/10.1016/j.ymben.2018.02.001
34. Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614
35. de Groot, D. H., Lischke, J., Muolo, R., Planqué, R., Bruggeman, F. J., & Teusink, B. (2020). The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cellular and Molecular Life Sciences*, *77*(3), 441–453. https://doi.org/10.1007/s00018-019-03380-2
36. Burgard, A. P., Nikolaev, E. V., Schilling, C. H., & Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, *14*(2), 301–312. https://doi.org/10.1101/gr.1926504
37. Larhlimi, A., David, L., Selbig, J., & Bockmayr, A. (2012). F2C2: A fast tool for the computation of flux coupling in genome-scale metabolic networks. *Bmc Bioinformatics*, *13*, 57. https://doi.org/10.1186/1471-2105-13-57
38. Reimers, A. C., Goldstein, Y., & Bockmayr, A. (2015). Generic flux coupling analysis. *Mathematical Biosciences*, *262*, 28–35. https://doi.org/10.1016/j.mbs.2015.01.003
39. Patil, K. R., Rocha, I., Förster, J., & Nielsen, J. (2005). Evolutionary programming as a platform for in silico metabolic engineering. *Bmc Bioinformatics*, *6*, 308. https://doi.org/10.1186/1471-2105-6-308
40. Schneider, P., & Klamt, S. (2019). Characterizing and ranking computed metabolic engineering strategies. *Bioinformatics*, *35*(17), 3063–3072. https://doi.org/10.1093/bioinformatics/bty1065
41. Vieira, V., Maia, P., Rocha, M., & Rocha, I. (2019). Comparison of pathway analysis and constraint-based methods for cell factory design. *Bmc Bioinformatics*, *20*(1), 350. https://doi.org/10.1186/s12859-019-2934-y
42. Hädicke, O., & Klamt, S. (2010). CASOP: A computational approach for strain optimization aiming at high productivity. *Journal of Biotechnology*, *147*(2), 88–101. https://doi.org/10.1016/j.jbiotec.2010.03.006
43. Klamt, S., Mahadevan, R., & Hädicke, O. (2018). When do two-stage processes outperform one-stage processes? *Biotechnology Journal*, *13*(2), 1700539. https://doi.org/10.1002/biot.201700539
44. Klamt, S. (2006). Generalized concept of minimal cut sets in biochemical networks. *Bio Systems*, *83*(2), 233–247. https://doi.org/10.1016/j.biosystems.2005.04.009
45. Klamt, S., Mahadevan, R., & von Kamp, A. (2020). Speeding up the core algorithm for the dual calculation of minimal cut sets in large metabolic networks. *Bmc Bioinformatics*, *21*(1), 510. https://doi.org/10.1186/s12859-020-03837-3
46. Klamt, S., & Gilles, E. D. (2004). Minimal cut sets in biochemical reaction networks. *Bioinformatics*, *20*(2), 226–234. https://doi.org/10.1093/bioinformatics/btg395
47. Mahadevan, R., von Kamp, A., & Klamt, S. (2015). Genome-scale strain designs based on regulatory minimal cut sets. *Bioinformatics*, *31*(17), 2844–2851. https://doi.org/10.1093/bioinformatics/btv217
48. Schneider, P., Kamp, A., & Klamt, S. (2020). An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets. *Plos Computational Biology*, *16*(7), e1008110. https://doi.org/10.1371/journal.pcbi.1008110
49. Klamt, S., Saez-Rodriguez, J., & Gilles, E. D. (2007). Structural and functional analysis of cellular networks with CellNetAnalyzer. *Bmc Systems Biology*, *1*, 2. https://doi.org/10.1186/1752-0509-1-2
50. Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., Feist, A. M., & Palsson, B. O. (2017). iML1515, a knowledgebase that computes Escherichia coli traits. *Nature Biotechnology*, *35*(10), 904–908. https://doi.org/10.1038/nbt.3956
51. Hädicke, O., & Klamt, S. (2017). *EColiCore2*: A reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Scientific Reports*, *7*, 39647. https://doi.org/10.1038/srep39647
52. Shabestary, K., & Hudson, E. P. (2016). Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis. *Metabolic Engineering Communications*, *3*, 216–226. https://doi.org/10.1016/j.meteno.2016.07.003
53. Navas, M. A., Cerdán, S., & Gancedo, J. M. (1993). Futile cycles in Saccharomyces cerevisiae strains expressing the gluconeogenic enzymes during growth on glucose. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(4), 1290–1294. https://doi.org/10.1073/pnas.90.4.1290
54. Hädicke, O., & Klamt, S. (2015). Manipulation of the ATP pool as a tool for metabolic engineering. *Biochemical Society Transactions*, *43*(6), 1140–1145. https://doi.org/10.1042/BST20150141
55. Hädicke, O., Bettenbrock, K., & Klamt, S. (2015). Enforced ATP futile cycling increases specific productivity and yield of anaerobic lactate production in Escherichia coli. *Biotechnology and Bioengineering*, *112*(10), 2195–2199. https://doi.org/10.1002/bit.25623
56. Liu, J., Kandasamy, V., Würtz, A., Jensen, P. R., & Solem, C. (2016). Stimulation of acetoin production in metabolically engineered Lactococcus lactis by increasing ATP demand. *Applied Microbiology and Biotechnology*, *100*(22), 9509–9517. https://doi.org/10.1007/s00253-016-7687-1
57. Semkiv, M. V., Dmytruk, K. V., Abbas, C. A., & Sibirny, A. A. (2016). Activation of futile cycles as an approach to increase ethanol yield during glucose fermentation in Saccharomyces cerevisiae. *Bioengineered*, *7*(2), 106–111. https://doi.org/10.1080/21655979.2016.1148223
58. Boecker, S., Zahoor, A., Schramm, T., Link, H., & Klamt, S. (2019). Broadening the scope of enforced ATP wasting as a tool for metabolic engineering in Escherichia coli. *Biotechnology Journal*, *14*(9), 1800438. https://doi.org/10.1002/biot.201800438
59. Boecker, S., Harder, B.-J., Kutscha, R., Pflügl, S., & Klamt, S. (2021). Increasing ATP turnover boosts productivity of 2,3-butanediol synthesis in Escherichia coli. *Microbial Cell Factories*, *20*(1), 63. https://doi.org/10.1186/s12934-021-01554-x

60. Zahoor, A., Messerschmidt, K., Boecker, S., & Klamt, S. (2020). ATPase-based implementation of enforced ATP wasting in Saccharomyces cerevisiae for improved ethanol production. *Biotechnology for Biofuels*, *13*(1), 185. https://doi.org/10.1186/s13068-020-01822-9
61. Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D., & Jensen, P. R. (2002). The glycolytic flux in Escherichia coli is controlled by the demand for ATP. *Journal of Bacteriology*, *184*(14), 3909–3916. https://doi.org/10.1128/jb.184.14.3909-3916.2002
62. Jouhten, P., Huerta-Cepas, J., Bork, P., & Patil, K. R. (2017). Metabolic anchor reactions for robust biorefining. *Metabolic Engineering*, *40*, 1–4. https://doi.org/10.1016/j.ymben.2017.02.010

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Schneider, P., Mahadevan, R., & Klamt, S. (2021). Systematizing the different notions of growth-coupled product synthesis and a single framework for computing corresponding strain designs. *Biotechnol. J.*, *16*, e2100236. https://doi.org/10.1002/biot.202100236