

Systems biology

Underground metabolism as a rich reservoir for pathway engineering

Szabolcs Cselgő Kovács^{1,2,†}, Balázs Szappanos^{1,2,3,*,†}, Roland Tengölics^{1,2},
Richard A. Notebaart ^{4,*} and Balázs Papp^{1,2,*}

¹HCEMM-BRC Metabolic Systems Biology Lab, Szeged, Hungary, ²Biological Research Centre, Institute of Biochemistry, Synthetic and Systems Biology Unit, Eötvös Loránd Research Network (ELKH), Szeged, Hungary, ³Department of Biotechnology, University of Szeged, Szeged, Hungary and ⁴Food Microbiology, Wageningen University & Research, 6700 HB Wageningen, The Netherlands

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Associate Editor: Pier Luigi Martelli

Received on January 14, 2022; revised on April 12, 2022; editorial decision on April 12, 2022; accepted on April 14, 2022

Abstract

Motivation: Bioproduction of value-added compounds is frequently achieved by utilizing enzymes from other species. However, expression of such heterologous enzymes can be detrimental due to unexpected interactions within the host cell. Recently, an alternative strategy emerged, which relies on recruiting side activities of host enzymes to establish new biosynthetic pathways. Although such low-level ‘underground’ enzyme activities are prevalent, it remains poorly explored whether they may serve as an important reservoir for pathway engineering.

Results: Here, we use genome-scale modeling to estimate the theoretical potential of underground reactions for engineering novel biosynthetic pathways in *Escherichia coli*. We found that biochemical reactions contributed by underground enzyme activities often enhance the *in silico* production of compounds with industrial importance, including several cases where underground activities are indispensable for production. Most of these new capabilities can be achieved by the addition of one or two underground reactions to the native network, suggesting that only a few side activities need to be enhanced during implementation. Remarkably, we find that the contribution of underground reactions to the production of value-added compounds is comparable to that of heterologous reactions, underscoring their biotechnological potential. Taken together, our genome-wide study demonstrates that exploiting underground enzyme activities could be a promising addition to the toolbox of industrial strain development.

Availability and implementation: The data and scripts underlying this article are available on GitHub at <https://github.com/pappb/Kovacs-et-al-Underground-metabolism>.

Contact: szappanos.balazs@brc.hu or richard.notebaart@wur.nl or papp.balazs@brc.hu

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Due to our strong dependency on fossil materials, there is a growing interest for the development of sustainable ways to replace them and to provide alternative approaches to produce chemical compounds with application in fields such as medicine (pharmaceuticals), cosmetics, materials (e.g. bioplastics) and food. One promising approach is the use of microbial cell factories to produce such value-added compounds. More specifically, the metabolic network of these species can be redesigned and engineered to optimize the production of the desired compounds (Hadadi and Hatzimanikatis, 2015; Ko *et al.*, 2020; Nielsen and Keasling, 2016; Wang *et al.*, 2017; Yim *et al.*, 2011).

Several metabolic engineering approaches have been developed to improve the production of value-added compounds (Okano *et al.*, 2018; Pontrelli *et al.*, 2018). Up till now, the most widely applied approach is the introduction and expression of heterologous metabolic reactions (i.e. reactions from a different organism), which generate heterologous biosynthetic pathways that enable the production of non-native value-added compounds by the host organism (Lechner *et al.*, 2016; Pickens *et al.*, 2011). Several computational approaches have been proposed to facilitate pathway engineering through heterologous reactions using genome-scale metabolic modeling (Ko *et al.*, 2020; Wang *et al.*, 2017). Genome-scale metabolic models are available for the most common microorganisms used in metabolic engineering and they can accurately predict how the

addition of heterologous enzymes affect metabolite production (Pharkya *et al.*, 2004). Although the vast repertoire of available heterologous reactions provide a huge potential for value-added compound production, there are several limitations. First, heterologous enzymes often require specific cofactors which cannot be provided by the host organism hindering the biosynthetic pathway of its proper working (Boynton *et al.*, 1996). Second, heterologous expression could lead to stress response due to protein overproduction or accumulation of toxic intermediates (Gill *et al.*, 2000; Martin *et al.*, 2003). Third, microorganisms containing heterologous reactions are considered as GMO, whose commercial application, especially in food industry, relies on the juristic regulations of countries.

One possible approach to overcome these limitations is to exploit what is known as the underground metabolism, i.e. the collection of enzyme side activities in a cell (D'Ari and Casadesús, 1998; Fig. 1). In addition to their native activities, most enzymes display weak side activities by which substrates are turned into products, albeit at low rates, due to the limited substrate specificity of the enzyme. Although the physiological effect of such underground reactions is mostly neglected given their inefficient kinetics, they can often be enhanced by only a few mutations (Aharoni *et al.*, 2005; Khersonsky and Tawfik, 2010; Notebaart *et al.*, 2014). Such genetic changes allow underground metabolism to contribute to microbial growth and adaptation to novel nutrient conditions (Cam *et al.*, 2016; King *et al.*, 2017; Notebaart *et al.*, 2014; 2018). Although underground activities have been utilized for pathway development in case studies, no comprehensive work has yet explored the biotechnological potential of the underground metabolic network (Notebaart *et al.*, 2018; Rosenberg and Commichau, 2019). In particular, it remains unexplored whether biochemical reactions catalyzed by enzyme side activities are as likely to contribute to new biosynthetic pathways toward value-added compounds as those catalyzed by heterologous enzymes.

Here, we aim to systematically assess the extent to which underground metabolism can increase the production of industrially important compounds. As genome-scale computational modeling of underground metabolism successfully predicts the potential to utilize new nutrient sources in *Escherichia coli* (Notebaart *et al.*, 2014), we reasoned that a similar approach could be employed to characterize the theoretical potential of underground reactions to produce industrially relevant chemical compounds. To this end, we integrated experimentally reported and predicted underground reactions into a genome-scale metabolic reconstruction of *E. coli* (Orth *et al.*, 2011) and characterized its biosynthetic properties across 64 different industrially important compounds spanning diverse applications, such as health, material sciences, chemical industry and food. Our computational analyses demonstrate that underground enzyme activities

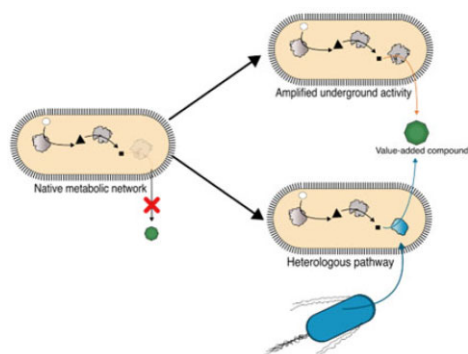


Fig. 1. Utilization of underground activities as an alternative to incorporating heterologous enzymes. The schematic figure depicts two strategies to enhance or establish the production of industrially relevant chemicals in microorganisms. The conventional strategy is to introduce one or more heterologous enzymes from a different organism (cell with long flagella, blue) to create a heterologous pathway capable of producing the value-added compound (lower part). A less well explored alternative is to amplify existing low-activity underground reactions (upper part, orange arrow) catalyzed by the endogenous enzymes of the organism (A color version of this figure appears in the online version of this article)

frequently enhance the production of value-added compounds with industrial importance, which is comparable to heterologous enzyme activities that enhance such value-added compounds. Notably, we find that activation of only a single underground reaction can already substantially increase the production of important industrial compounds, such as a precursor of bioplastics (3-hydroxypropionate) and a potential biofuel (1-butanol). Overall, our work reveals that the underground metabolic network provides a promising metabolic engineering tool to improve the production of both native and non-native value-added compounds.

2 Materials and methods

2.1 Reconstruction of the *E. coli* underground metabolism version 2.0

To study the industrial potential of the underground metabolic network, we extended a previously reconstructed underground metabolic network of *E. coli* K-12 MG1655 (Notebaart *et al.*, 2014) with additional underground reactions. These additional underground reactions were *in silico* predicted by the PROPER algorithm and 20% of them were validated experimentally (Oberhardt *et al.*, 2016). In brief, these underground activities were predicted to be associated with an enzyme based on its sequence similarity to orthologous enzymes (Oberhardt *et al.*, 2016). We next integrated the set of underground reactions into the native *E. coli* genome-scale metabolic model iJO1366 (Orth *et al.*, 2011). We removed duplicate reactions and ‘perpetuum mobile’ cycles, that is, flux distributions that are able to produce energy without the presence of available nutrients (Fritzemeier *et al.*, 2017). Note that underground reactions u0291 and u0227 were adjusted to contain 1-propanol instead of isopropanol according to the literature references in the Brenda database. The resulting metabolic model (hereby termed underground model) contains 3146 reactions and 2172 metabolites, of which 563 reactions and 367 metabolites are not part of the native model, respectively. The reconstruction is available as a computational SBML model (Supplementary File S1, also downloadable from <https://github.com/pappb/Kovacs-et-al-Underground-metabolism>).

2.2 Compiling a list of value-added compounds

To get a broad view of the potential of underground metabolism in industrial bioproduction, we compiled a comprehensive list of value-added compounds (Supplementary Dataset S1). A compound was considered to hold a value if it fulfills a role in industrial processes such as precursor, building block, reagent, solvent or final product. Most compounds are adapted from the Standard Industrial Classification Manual of United States Department of Labor (Industrial Organic Chemicals, Industry Group 286; Major Group 28: Chemicals and Allied Products | Occupational Safety and Health Administration). The other major source is an exhaustive report of Werpy and Petersen presenting the most wanted value-added compounds whose sustainable production is a crucial issue (Werpy and Petersen, 2004). In addition, the list was extended by individual studies that attempted bioproduction of value-added compounds (Supplementary Dataset S1). Because we are interested in compounds that are known as metabolites and because the association of compounds with their representation in the model is based on their KEGG database IDs (Kanehisa *et al.*, 2017), we excluded compounds without a KEGG ID. This resulted in a list of 160 value-added compounds that qualify as metabolites and used for subsequent analyses (see Supplementary Dataset S1).

2.3 Defining *in silico* media

We simulated the growth of *E. coli* in M9 minimal medium adapted from a glucose minimal medium of a previous model (Feist *et al.*, 2007). In order to estimate the influence of nutrients, we run our yield calculations on seven different carbon sources: D-glucose (Bhatia *et al.*, 2012), glycerol (Clomburg and Gonzalez, 2013), D-xylose (Bhatia *et al.*, 2012), D-fructose (Aristidou *et al.*, 1999), L-fucose (Kim *et al.*, 2019), L-arabinose (Bhatia *et al.*, 2012) and acetate (Wu *et al.*, 2016).

These carbon sources are either already used or suggested to be used in industrial compound production by prior works. In our simulations, we enabled the uptake of only a single carbon source at a time.

2.4 Calculating the production yield of value-added compounds

To calculate the production yield of value-added compounds we applied Flux Balance Analysis (FBA) (Orth et al., 2010). First, we constrained the biomass production to 10% of its maximal value to ensure that cellular growth is not completely halted by the compound production. Second, we extended the model with a virtual ‘demand’ reaction that has the value-added compound as the substrate and no product to mimic the industrial extraction of the compound. Third, we run FBA simulation to maximize the flux on the demand reaction, i.e. to maximize the production of the value-added compound. The flux of the demand reaction equals to the production rate of the target compound. Finally, we calculated the Maximum Theoretical Yield (shortened as yield) as a metric to assess the efficiency of predicted biosynthetic pathways according to the following equation (Campodonico et al., 2014):

$$\text{Yield} = \frac{\text{Production rate (Value - added compound)}}{\text{Uptake rate (Carbon source)}} \\ * \frac{\text{Carbon atom number (Value - added compound)}}{\text{Carbon atom number (Carbon source)}}$$

We defined biotechnologically relevant (i.e. significant) yield increments in the underground metabolic model as those that enhanced the theoretical yield of the native model by at least 5%. The same criterion was applied when determining the impact of adding heterologous reaction sets to the native model. Although this threshold is arbitrary, our results were rather insensitive to the actual value (see Section 3).

2.5 Determining the minimum number of underground reactions needed for increasing the production yield

We used a mixed integer linear programming (MILP) algorithm inspired by our previous study and the Optcouple algorithm to identify minimal underground reaction sets that are directly involved in the production of value-added compounds (Jensen et al., 2019; Szappanos et al., 2016). Our algorithm looked for a minimal set of underground reactions sufficient for yield increment for each target compound. Although there might be multiple such minimal reaction sets we collected only one for each target compound as an example reaction set and to determine the minimal number of underground reactions necessary for yield increment. We excluded those value-added compounds from the MILP analysis where the production yield is enhanced by less than 5% by underground reactions. The basis of the MILP problem was the steady-state assumption:

$$Sv = 0$$

where S is the stoichiometric matrix and v is the flux vector for all reactions. The reactions of the model were distinguished according to their origin: native or underground. The flux constraints on the *E. coli* reactions were the same as in FBA:

$$LB_i \leq v_i \leq UB_i$$

where v_i is the flux of a reaction, LB_i and UB_i is the lower and upper flux bounds, respectively.

We interpreted our goal as a dual optimization problem. While the primal objective function was the flux maximization of a value-added compound, the dual objective was the minimization of the number of active underground reactions. The activation of underground reactions was controlled with binary variables:

$$b_i \in B \forall i \in \{1, \dots, N'\}$$

where b is a binary variable, i is the index of an underground reaction, B is the set of binary variables and N' is the number of

underground reactions. The binary variable b_i tells whether the underground reaction r'_i ($i = 1, \dots, N'$) is active ($b_i = 1$) or not ($b_i = 0$). The following equations ensure these rules:

$$v'_i \geq \varepsilon * b_i \\ UB'_i * b_i \geq v'_i$$

where v'_i is the flux and UB'_i is the maximal possible flux of underground reaction r'_i , while ε is the minimal non-zero flux value (in our calculations $\varepsilon = 10^{-8}$). Reversible underground reactions of the underground network were decomposed into two opposing irreversible reactions. This way the fluxes of the underground reactions can only take positive values, which is a prerequisite for MILP. In addition, to avoid having two opposing reactions derived from the same reversible reaction being active simultaneously we introduced the following constraint:

$$b_i + b_j \leq 1; (i, j) \in \{\text{set of opposing reaction pairs}\}$$

To solve the first optimization problem, we calculated the maximal production rate for each value-added compound (see Section 2.4). Next, for each value-added compound we constrained the flux of the demand reaction of the value-added compound to be equal to its maximum value to ensure the maximal production of the compound.

Finally, the second objective of the MILP problem was to minimize the active underground reactions:

$$\text{minimize } \sum b_i; i \in \{1, \dots, N'\}$$

The result of this optimization is the minimum number of underground reactions whose collective presence is required for the enhanced value-added compound production.

2.6 Statistical comparison of underground and heterologous reactions

To compare the industrial potentials of underground and heterologous enzyme activities, we extended the native *E. coli* metabolic model iJO1366 (Orth et al., 2011) with a large set of heterologous reactions (hereby termed heterologous model). Data on heterologous reactions were obtained from the MetaCyc database (version 22.5), a comprehensive database of metabolism from all domains of life (Caspi et al., 2020). All MetaCyc reactions not present in the *E. coli* native network were considered as heterologous. Next, we discarded heterologous reactions that were not mass-balanced or contained metabolites without a KEGG ID association (Kanehisa et al., 2017). We used the associated KEGG IDs to identify compounds that are already present in the native *E. coli* network. We also added the remaining non-native compounds to the metabolic model and integrated the heterologous reactions into the model. After that, we removed duplicate reactions and ‘perpetuum mobile’ cycles, that is, flux distributions that are able to produce energy without the presence of available nutrients (Fritzemeier et al., 2017). Finally, to make the heterologous model comparable to the underground model, we added further exchange and native reactions that were also incorporated into the underground model based on literature evidence (Notebaart et al., 2014). The resulting heterologous metabolic model contains 8050 reactions and 5538 metabolites, whereof 5686 reactions and 3495 metabolites are not part of the native model, respectively. The heterologous model is available as a computational Systems Biology Markup Language (SBML) model (Supplementary File S2, <https://github.com/pappb/Kovacs-et-al-Underground-metabolism>).

Since the set of collected heterologous reactions is much larger than the set of underground reactions, we generated 1000 heterologous subsets. Each subset contained a random set of heterologous reactions chosen from the heterologous model. The number of heterologous reactions in each random set matched the number of underground reactions in the underground model. For each of the 1000 heterologous subsets, we then calculated the maximum theoretical yield and yield increments of the value-added compounds in the same way as we did for the underground model (see Section 2.4).

Out of the 160 value-added compounds with KEGG IDs, 110 were present in the heterologous model either as native or heterologous metabolites and therefore the heterologous production of these 110 compounds were tested. Calculations with the heterologous subsets were done only on glucose minimal medium to reduce computational time.

The combined effect of underground and heterologous reactions was assessed using a model reconstruction that contain both the heterologous and underground reactions. This model is based on the heterologous model described above and extended with the underground reaction sets (the model is available as SBML file in [Supplementary File S3](#) and under the following GitHub link: <https://github.com/pappb/Kovacs-et-al-Underground-metabolism>). Yield increments were calculated for this combined model under glucose minimal medium as described in Section 2.4.

2.7 Software and computation used in metabolic simulations

All simulations were implemented in python3.6 (Python Software Foundation. Python Language Reference, version 3.6. Available at <http://www.python.org>) using cameo (Cardoso *et al.*, 2018) and cobrapy (Ebrahim *et al.*, 2013) python packages for constraint-based modeling. As an optimizer for linear programming and MILP we used GUROBI 8.1 (Gurobi Optimization, LLC, 2021). The linear programming was done on a 64-bit Ubuntu Linux system with an Intel Core-i7 quad core processor. MILP problems were solved on a Red Hat Enterprise Linux Server release 6.2 with 96 Intel Xeon central processing units.

3 Results

3.1 Reconstructing an expanded model of *E. coli* underground metabolism

To achieve high coverage of potential underground reactions of *E. coli*, we extended our previously published list of experimentally reported *E. coli* underground activities (Notebaart *et al.*, 2014) with predicted ones and incorporated them into the iJO1366 genome-scale metabolic reconstruction of this species (Orth *et al.*, 2011). Specifically, we utilized the predictions of the PROPER algorithm, which assigns native activities of homologous enzymes from other bacteria as underground activities to *E. coli* enzymes (Oberhardt *et al.*, 2016). Importantly, this algorithm achieved good overlap with multicopy suppression studies in *E. coli* where the over-expression of a ‘replacer’ gene rescues an otherwise lethal loss-of-function mutation.

Our updated underground network contains 543 underground reactions, which is approximately 20% of the number of native reactions in the genome-scale model. By incorporating the underground reactions into the *E. coli* metabolic network, we also introduced 311 novel compounds that are not present in the native network. Hence, the extended metabolic network model contains 22% more metabolites than the native model (Fig. 2, [Supplementary Dataset S1](#)). Overall, the updated underground metabolic network reconstruction contains 107% and 12% more reactions and metabolites, respectively, compared to the previous version, indicating a substantial increase in coverage.

Underground reactions do not necessarily have the potential to carry flux due to missing connections to the rest of the metabolic network. Nevertheless, we found that 79% of the compiled underground reactions are either fully or partially connected to the native network (i.e. either the substrates or the products, or both are present in the native network). This result suggests that many underground reactions may have the potential to participate in novel biosynthetic pathways.

3.2 Underground reactions can often increase the yield of value-added compounds

To explore the potential utility of underground reactions for bioproduction of desired chemicals, we compiled a list of value-added

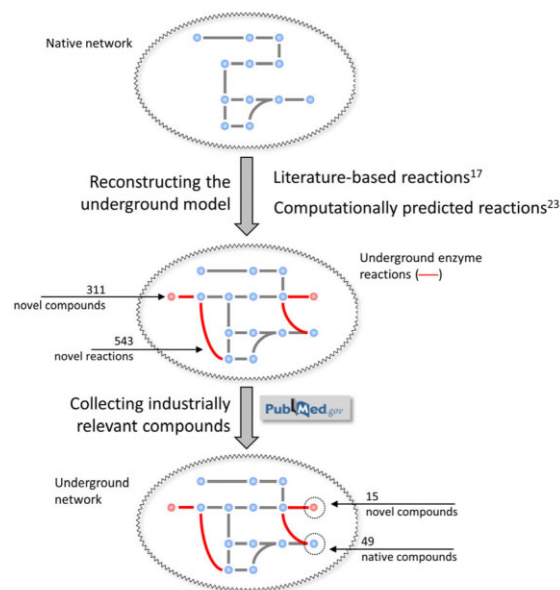


Fig. 2. Reconstructing an expanded underground metabolic network of *E. coli*. First, the native *E. coli* metabolic model was extended with both experimentally validated (Notebaart *et al.*, 2014) and predicted (Oberhardt *et al.*, 2016) underground reactions. Next, value-added compounds were collected from the literature and those that are present in the model were kept

compounds from literature. As we are interested in bioproduction, we focused on compounds that have been described as metabolites according to the KEGG database (Kanehisa *et al.*, 2017). Our list includes not only the desired end products (e.g. biofuels) but also precursors, building blocks and relevant reagents and solvents as well (see Section 2). Out of the 160 value-added metabolites collected, 64 (40%) are present in the extended metabolic network reconstruction and thus have the potential to be produced by *E. coli*. We excluded 20 compounds because they cannot be produced in any of the simulated media even when all underground reactions are available for biosynthesis. The remaining set of 44 value-added compounds consists of 42 native metabolites that are already present in the native *E. coli* network and 2 novel metabolites that are participating in underground reactions only. Importantly, 16 out of the 44 target-compounds are among the top 30 most sought-after value-added compounds to be produced from renewable carbon sources (Werpy and Petersen, 2004). For example, ethylene glycol is used as antifreeze and also as a building block for plastics (Curme Jr and Young, 1925; Liu *et al.*, 2013).

Underground reactions may contribute to the production of value-added compounds in two ways. First, they may open up more efficient biosynthetic pathways for a compound that can already be produced by the native network. Second, they may enable the production of entirely new compounds that cannot be produced by the native network. To investigate the feasibility of these two scenarios, we systematically tested whether the presence of underground reactions increases the maximal theoretical yield of each value-added compound using flux balance analysis (Campodonico *et al.*, 2014) (see Section 2). Maximum theoretical yield (yield for short) measures production efficiency by calculating the fraction of carbon atoms coming from the carbon source that is converted into the production of the target compound (Campodonico *et al.*, 2014).

We allowed the utilization of all underground reactions simultaneously to identify all potential cases where the underground network facilitates the production of a target compound, including those where multiple underground reactions are required. These simulations were run on glucose as the sole carbon source. We report that underground reactions enhanced the maximum theoretical yield by more than 5% in 9 out of the 44 target compounds (Fig. 3A). These 9 compounds have various industrial applications including plastic manufacturing, flavoring, antifreeze production,

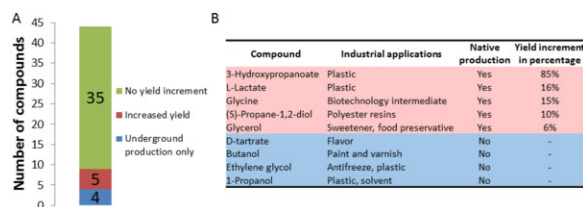


Fig. 3. Impact of underground metabolism on the production potential of value-added compounds. The impact of underground reactions is inferred by simultaneously enabling all underground reactions to allow for potential pathways that require multiple underground reactions. (A) Number of value-added compounds showing no yield increment (35, green), showing increased yield in the underground model compared to the native model (5, red) and showing production only in the presence of underground reactions (4, blue). (B) Table of compounds with increased yield (native production available, red) and exclusive production (no native production, blue) in the presence of underground reactions. Yield increment percentage cannot be calculated for the latter as the native model cannot produce those compounds (A color version of this figure appears in the online version of this article)

etc. (Fig. 3B). Increasing the threshold of yield increment to 10% would eliminate only a single hit (glycerol), showing that our results are robust to this parameter. In more than half of the cases (five out of nine), the native metabolic network is already capable of producing the compound but the underground reactions opened a new pathway with a higher yield. Out of those four compounds not produced by the native model (see Fig. 3B), two (D-tartrate and butanol) were already present in the native network, indicating that the underground reactions created new routes between existing parts of the native network to enable their production. In the remaining two cases (ethylene glycol and 1-propanol), the underground reactions extended the native network to reach the target compound (Fig. 4C).

To assess the extent to which these results depend on the applied nutrient conditions, we repeated the simulations on seven other carbon sources deemed relevant in metabolic engineering (see Materials and Methods). Generally, the yield increments showed little variation between carbon sources indicating that our results are robust to the choice of carbon source (Supplementary Dataset S1). This general lack of condition-dependency can be explained by the fact that most target compounds are biosynthesized from intermediates of central carbon metabolism with the sole exception of acrolein. Notably, acrolein can be produced only on D-xylose as the sole carbon source (see Supplementary Dataset S1).

Collectively, these results show that underground reactions often have the potential to increase the production efficiency of value-added compounds and can even confer the ability to synthesize new chemicals.

3.3 Yield increase can be achieved with a handful of underground reactions

Biosynthetic pathways with fewer underground reactions are expected to be easier to implement in biotechnological applications. Therefore, we next examined the number of underground reactions that are directly involved in the biosynthesis of a given target compound. This was achieved by applying a mixed integer linear programming (MILP) algorithm to find the minimum number of underground reactions necessary for the yield increment (Fig. 4, see Materials and Methods). In two-thirds of the cases (six out of nine), the presence of a single underground reaction is sufficient to increase the production yield and a maximum of four underground reactions are sufficient in all cases (Fig. 4A and B, Supplementary Dataset S1). The set of underground reactions involved in the production of specific target compounds are generally unique, with only a single underground reaction, catalyzed by glycerol dehydrogenase (*gldA*), being advantageous for the production of more than one compound.

Notably, in three out of the four target compounds that cannot be produced by the native model, the involved underground reactions are located at the end of the biosynthetic pathway (Supplementary Dataset S1, Fig. 4C). In contrast, target compounds that can already

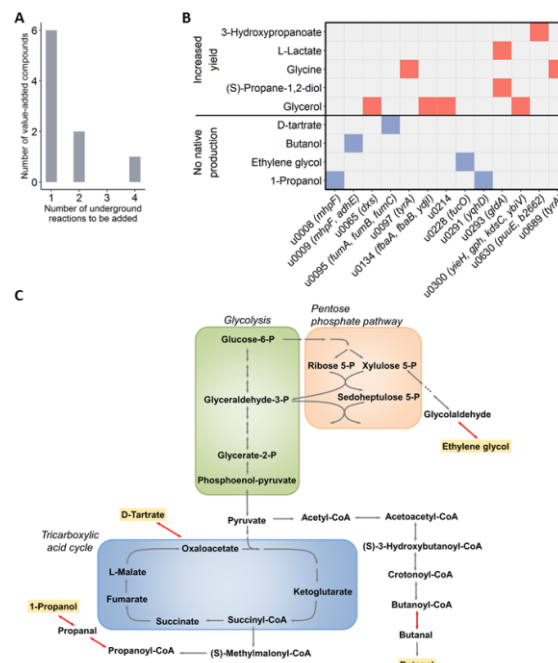


Fig. 4. The presence of one or two underground reactions can be sufficient to increase production yield. (A) Distribution of the number of underground reactions needed to increase production yields of target compounds. No more than four underground reactions have to be added to the native model to increase the yield in all cases. Furthermore, in most cases a single underground reaction is sufficient. (B) Matrix of underground reactions that contribute to the yield increments of each target compound. The columns represent the underground reactions (e.g. u0008, etc.) and their standard gene associations in parentheses, while the rows represent the target compounds. Cells in the upper part (marked as 'Increased yield', red) belong to metabolites that can also be produced by the native metabolic network while the ones in the lower part (marked as 'No native production', blue) require the presence of underground reactions. Multiple cells in one row account for multiple underground reactions whose joint presence is needed for production of the target compound. (C) Schematic view of central carbon metabolism (gray arrows) with underground reactions (red arrows) that enable the production of new compounds (yellow highlight) that cannot be produced by the native metabolic network. For details, see Supplementary Dataset S1 (A color version of this figure appears in the online version of this article)

be produced by the native model are never direct products of the underground reactions (Supplementary Dataset S1). These results suggest that underground reactions directly involved in the production of the target compound have higher impact on the yield. Future studies are demanded to test the generality of this hypothesis.

3.4 Empirical support for the biotechnological application of underground reactions

Literature survey provides empirical support for several of our computational predictions. First, consistent with the predictions, overexpressing *gldA* is a crucial part of the experimental strain design for (S)-propane-1,2-diol production (Clomburg and Gonzalez, 2011). The gene *gldA* is an endogenous aldehyde reductase with a primary role in removal of dihydroxyacetone by converting it to glycerol (Subedi et al., 2008). However, it has also been shown to catalyze the conversion of methylglyoxal to lactaldehyde, a precursor of (S)-propane-1,2-diol (Clomburg and Gonzalez, 2011). Our method successfully identified this reaction as the underground reaction necessary to enhance (S)-propane-1,2-diol production. Furthermore, in line with our simulations, experiments show that production of (S)-propane-1,2-diol is more efficient on glycerol than on glucose as the carbon source (Clomburg and Gonzalez, 2011) (see Supplementary Dataset S1). Another example is the production of ethylene glycol, where a previous experimental implementation utilized the same underground activity and gene (*fucO*) as predicted here. Specifically, a key step in engineering a novel pathway for ethylene glycol

production relied on the underground activity of FucO, which converts glycolaldehyde to ethylene glycol (Pereira *et al.*, 2016).

1-Propanol emerged as one of the compounds that cannot be produced by native *E. coli* metabolism, but can be synthesized by simultaneously adding two underground reactions to the network (Fig. 4B). 1-Propanol is used as a solvent during production of cosmetics, and it is also used to manufacture cellulose-based plastics (Gonzalez-Garcia *et al.*, 2017). Moreover, 1-propanol is a promising alternative biofuel to bioethanol considering its higher energy density (Jun Choi *et al.*, 2012). There are multiple pathways engineered to produce 1-propanol and one of them includes the two underground reactions predicted by our method (Jun Choi *et al.*, 2012) (Fig. 4B). In particular, the study of Jun Choi *et al.* (2012) demonstrates the experimental feasibility of the predicted underground pathway converting propionyl-CoA into propionyl-aldehyde and further to 1-propanol. Note, however, that the associated gene (*adhE*) in this experimental study is different from those predicted by our analysis (*mhpF* and *yqhD*), indicating that the underground network reconstruction is far from complete.

Overall, despite the sparse usage of enzyme side activities in metabolic engineering (Pontrelli *et al.*, 2018), these examples provides support for our simulations.

3.5 Underground reactions show a similar potential to produce value-added compounds as heterologous reactions

To systematically assess the utility of underground pathways in the production of value-added compounds, we next compared their ability to increase production yield to that of heterologous pathways from other species, which are commonly employed in strain development. To this end, we first compiled a dataset of heterologous enzymatic reactions that are absent from *E. coli*, but have been described in other organisms according to the MetaCyc database (see Section 2). This resulted in 5686 distinct heterologous reactions that were added to the native *E. coli* model.

To compare the production potentials of underground and heterologous reactions while controlling for the different sizes of these two sets, we generated random sets of heterologous reactions having the same size as the total number of underground reactions. Next, we calculated how frequently these subsets of ‘heterologous’ reactions facilitate the production of value-added compounds compared to the model containing underground reactions. Note that out of the 160 compiled value-added metabolites, here we evaluated all 110 ones that participate in either a heterologous or a native reaction (Section 2). We found that, on average, random heterologous reaction sets increase the production of 6.1 target compounds. Statistical comparison to the nine cases where the underground reactions increased the production yield of a target compound revealed no significant difference (randomization test; $P=0.168$; Fig. 5A). Furthermore, underground reactions can produce a similar number of novel compounds which cannot be produced by the native network (4), compared to heterologous reaction sets (2.4 on average; randomization test; $P=0.238$; Fig. 5B). Together, these results show that underground reactions have a comparable potential to contribute to the production of value-added compounds as heterologous reactions.

Last, we asked whether a combined strategy using both underground and heterologous activities would improve the production of specific compounds to extents that cannot be achieved by either reaction repertoire alone (i.e. synergistic effects). To examine this, we calculated the yield increment using a model containing all underground and heterologous reactions simultaneously. Notably, we identified four value-added compounds that showed an enhanced production yield only when both underground and heterologous reactions were made available for the calculations (see Supplementary Dataset S1). This result shows that the combined use of underground and heterologous enzyme activities can further extend the biotechnological potential of an organism.

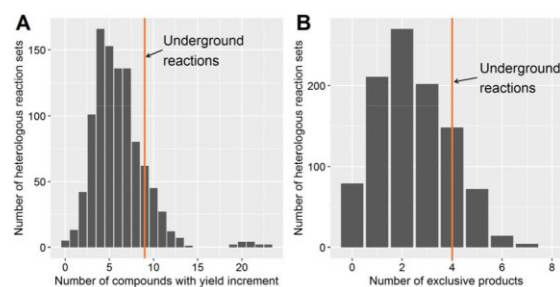


Fig. 5. Underground reactions show similar potential to produce value-added compounds as heterologous reactions. The figure compares the impact of adding the set of 543 underground reactions to the native network (vertical line) with that of adding the same number of randomly selected heterologous reactions (histograms). (A) Distribution of the number of target compounds where sufficient yield increment was attained across 1000 instances of randomly selected heterologous reaction sets. Underground reactions show comparable yield increment (nine cases, $P=0.168$, randomization test). (B) Distribution of the number of non-native target compounds which cannot be produced by the native metabolic network. There is no significant difference compared to underground reactions (four cases, $P=0.238$, randomization test). Vertical line represents the production properties of the underground metabolic network on both plots

4 Discussion

By extending the native metabolic network of *E. coli* by underground reactions, we present a framework to further broaden the metabolic engineering application for the production of value-added compounds (Pontrelli *et al.*, 2018). Our genome-scale computational analysis gave several new insights. First, we show that underground metabolism can enhance the production yield of numerous industrially relevant compounds and even enable the production of new compounds that cannot be produced by the native metabolic network.

Second, we also demonstrate that production of a given value-added compound through underground metabolism often hinges on one or few underground reactions only. This implies that it would be sufficient to engineer only few enzyme side activities for any given target compound. Engineering a small number of enzymatic steps would likely benefit a successful metabolic engineering strategy, especially since it concerns weak-side activities of the enzyme. There are several ways to translate our predictions toward the production of engineered strains *in vivo*, which is clearly the next major step. First, mutations that increase the underground activity could be engineered via various genome editing techniques, such as CRISPR-Cas and MAGE-based methods (Csörgő *et al.*, 2020; Jakočiūnas *et al.*, 2016; Wang *et al.*, 2009). Second, adaptive laboratory evolution has shown to be successful in increasing underground activities (Guzmán *et al.*, 2019; Pontrelli *et al.*, 2018), and third, a combination of editing and evolution could be applied (Pontrelli *et al.*, 2018; Wannier *et al.*, 2020).

Perhaps the most important new insight provided by our study is that underground and heterologous reactions have similar theoretical potentials to contribute to the production of value-added compounds. Thus, underground enzyme activities may provide a complementary source of biochemical reactions for overproduction purposes. This is a notable result, because the use of heterologous genes might be unfavorable for applications that demand a GMO-free status, such as food fermentation products. In contrast, the use of underground reactions coupled with adaptive evolution does not involve the introduction of specific DNA from other organisms. As such, the second approach may contribute to applications that go beyond GMO. Moreover, in a preliminary analysis we found several cases where specific underground and heterologous activities are jointly required to improve the production yield of value-added compounds. This result suggests that combining the two reaction repertoires can be advantageous for industrial applications. Clearly, future works are needed to fully explore this possibility. Last, we expect that underground activities might have a limited potential compared to heterologous enzyme activities in one particular area of

application: many industrially relevant metabolites are produced through secondary metabolism in plants, which are unlikely to be producible by side activities of microbial host enzymes.

We note that our estimate of the industrial potential of underground reactions might be distorted by at least two phenomena. On the one hand, the applied modeling framework might overestimate the contribution of underground reactions to new pathways for several reasons. First, we used a simple FBA framework that ignores thermodynamic realizability and therefore some of the predicted pathways might be unrealistic under physiologically relevant metabolite concentrations. Genome-scale modeling methods that incorporate thermodynamics as well as enzyme constraints through kinetics could address this shortcoming (Hoppe et al., 2007; Salvy et al., 2019; Sánchez et al., 2017). Recent frameworks allow the incorporation of enzyme kinetics data as well as quantitative omics data into models, which could result in the prediction of biological relevant phenotypes through FBA (Filippo et al., 2022; Sánchez et al., 2017). Second, underground enzyme activities should not interfere with the native metabolic network structure to become biochemically and physiologically functional. As such, it might be challenging to enhance underground activities *in vivo*, especially if extensive protein engineering is needed (Porokhin et al., 2021). Third, the FBA-based modeling framework might be over-optimistic and in reality, multiple genetic modifications are needed to achieve the desired production. For example, a single underground activity associated with the *fucO* gene is sufficient for efficient ethylene glycol production *in silico*. However, in addition to *fucO* overexpression, three other native and two heterologous enzymes had to be overexpressed and a gene deleted to achieve a high-flux pathway from glucose to ethylene glycol *in vivo* (Pereira et al., 2016). Similarly, to produce (S)-Propane-1,2-diol, a previous work overexpressed three genes, replaced an enzyme with a more efficient heterologous enzyme and disrupted pathways that alter the flux from the synthesis (Clomburg and Gonzalez, 2011). In contrast, the addition of a single underground activity is sufficient to increase the predicted yield in our simulations. Therefore, hits from our simulations are potentially needed to be expanded with further pathway improvements, including deletion of genes that divert the flux from biosynthesis and modifications that alter the redox balance. Nevertheless, our method gives suggestions that open new ways to produce important chemicals after further refinement. On the other hand, our knowledge of underground activities is still rudimentary and there might be orders of magnitude more side activities than currently known that could potentially be recruited for new pathways. Notably, there are various promising recent reports to predict metabolic (side) reactions from cheminformatics, enzyme structures and machine learning (Amin et al., 2019; Carbonell et al., 2014; Carbonell and Faulon, 2010; Koch et al., 2017; Mou et al., 2021; Robinson et al., 2020). We therefore anticipate that future advances in machine learning and cheminformatics will further expand the known ‘metabolic reaction space’ of species, i.e. the total number of metabolic reactions that could potentially be active in a species (Hafner et al., 2021; Tyzack et al., 2019). This ‘space’ could be exploited for the production of value-added compounds (Campodonico et al., 2014; Carbonell et al., 2014).

The present study focused on underground reactions as raw materials for building new pathways, however, promiscuous activities of enzymes in the existing metabolic network may also affect negatively the production of target compounds (Kim and Copley, 2012). For example, it has been shown that promiscuous phosphatase activities redirect flux from a heterologous terpenoid biosynthetic pathway, hence decreasing its efficiency (Wang et al., 2018). A more complete knowledge of the repertoire of underground reactions, including those catalyzed by native and heterologous enzymes alike, would therefore be instrumental to avoid network disruptions arising from enzyme promiscuity (Porokhin et al., 2021).

Our research focuses on public information of value-added compounds used in the industry, but this is likely an underestimate of the total complement of compounds of interest. Therefore, our approach could easily be adapted for the needs of the industry to incorporate their compound of interest. Moreover, we report that

changes in the exact nutrient environment does not alter the predictions substantially, since the value-added compounds are produced from central metabolism. This may, however, change once the underground metabolic network is further extended in the future, as well as when additional value-added compounds are added to the network. Hence, our approach may also predict new industrially relevant cost-reducing environments to produce similar product yield or even increased yields. Our results pave the way for exploiting underground metabolism to produce novel strains and we anticipate that a growing interest in underground metabolism will go together with a rise of biotechnological applications.

Acknowledgements

The authors thank Dávid Balogh, Bálint Kintsés and Ákos Nyerges for helpful discussions.

Funding

This work was supported by the following research grants: National Research, Development and Innovation Office [Élvonal Program KKP 129814 and KH 125616 to B.P.]; the “Lendület” program of the Hungarian Academy of Sciences [LP2009-013/2012 to B.P.]; GINOP-2.3.2-15-2016-00026 (iChamber, B.P.); The European Union’s Horizon 2020 Research and Innovation Program [739593 to B.P.]; The National Laboratory of Biotechnology Grant [NKFIH-871-3/2020 to B.P.]; The Hungarian Academy of Sciences Premium Postdoctoral Research Program [PREMIUM-2018-294 to B.S.]; Janos Bolyai Research Fellowship from the Hungarian Academy of Sciences [BO/00728/21/8 to B.S.]; New National Excellence Program of the Ministry of Human Capacities [Bolyai+, UNKP-21-5-SZTE-564 to B.S.]; and the National Research, Development and Innovation Office [PD 128271 to R.T.].

Conflict of Interest: none declared.

References

- Aharoni, A. et al. (2005) The ‘evolvability’ of promiscuous protein functions. *Nat. Genet.*, **37**, 73–76.
- Amin, S.A. et al. (2019) Towards creating an extended metabolic model (EMM) for *E. coli* using enzyme promiscuity prediction and metabolomics data. *Microb. Cell Factories*, **18**, 109.
- Aristidou, A.A. et al. (1999) Improvement of biomass yield and recombinant gene expression in *Escherichia coli* by using fructose as the primary carbon source. *Biotechnol. Prog.*, **15**, 140–145.
- Bhatia, L. et al. (2012) An economic and ecological perspective of ethanol production from renewable agro waste: a review. *AMB Express*, **2**, 65.
- Boynton, Z.L. et al. (1996) Cloning, sequencing, and expression of clustered genes encoding beta-hydroxybutyryl-coenzyme A (CoA) dehydrogenase, crotonase, and butyryl-CoA dehydrogenase from *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.*, **178**, 3015–3024.
- Cam, Y. et al. (2016) Engineering of a synthetic metabolic pathway for the assimilation of (D)-xylose into value-added chemicals. *ACS Synth. Biol.*, **5**, 607–618.
- Campodonico, M.A. et al. (2014) Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path. *Metab. Eng.*, **25**, 140–158.
- Carbonell, P. et al. (2014) XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res.*, **42**, W389–W394.
- Carbonell, P. and Faulon, J.-L. (2010) Molecular signatures-based prediction of enzyme promiscuity. *Bioinf. Oxf. Engl.*, **26**, 2012–2019.
- Cardoso, J.G.R. et al. (2018) Cameo: a python library for computer aided metabolic engineering and optimization of cell factories. *ACS Synth. Biol.*, **7**, 1163–1166.
- Caspi, R. et al. (2020) The MetaCyc database of metabolic pathways and enzymes – a 2019 update. *Nucleic Acids Res.*, **48**, D445–D453.
- Clomburg, J.M. and Gonzalez, R. (2013) Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals. *Trends Biotechnol.*, **31**, 20–28.
- Clomburg, J.M. and Gonzalez, R. (2011) Metabolic engineering of *Escherichia coli* for the production of 1,2-propanediol from glycerol. *Biotechnol. Bioeng.*, **108**, 867–879.
- Csörgő, B. et al. (2020) Targeted mutagenesis of multiple chromosomal regions in microbes. *Curr. Opin. Microbiol.*, **57**, 22–30.

- Curme, G. Jr., and Young, C. (1925) Ethylene glycol. *Ind. Eng. Chem.*, **17**, 1117–1120.
- D'Ari, R., and Casadesús, J. (1998) Underground metabolism. *Bioessays*, **20**, 181–186.
- Ebrahim, A. et al. (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.*, **7**, 74.
- Feist, A.M. et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.
- Filippo, M.D. et al. (2022) INTEGRATE: model-based multi-omics data integration to characterize multi-level metabolic regulation. *PLoS Comput. Biol.*, **18**, e1009337.
- Fritzemeier, C.J. et al. (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal. *PLoS Comput. Biol.*, **13**, e1005494.
- Gill, R.T. et al. (2000) A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metab. Eng.*, **2**, 178–189.
- Gonzalez-Garcia, R.A. et al. (2017) Microbial propionic acid production. *Fermentation*, **3**, 21.
- Gurobi Optimization, LLC. (2021) *Gurobi Optimizer Reference Manual*. Gurobi Optimization, LLC., Beaverton, Oregon, USA.
- Guzmán, G.I. et al. (2019) Enzyme promiscuity shapes adaptation to novel growth substrates. *Mol. Syst. Biol.*, **15**, e8462.
- Hadadi, N., and Hatzimanikatis, V. (2015) Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.*, **28**, 99–104.
- Hafner, J. et al. (2021) A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives. *Nat. Commun.*, **12**, 1760.
- Hoppe, A. et al. (2007) Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst. Biol.*, **1**, 23.
- Jakočičinas, T. et al. (2016) CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.*, **34**, 44–59.
- Jensen, K. et al. (2019) OptCouple: joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metab. Eng. Commun.*, **8**, e00087.
- Jun Choi, Y. et al. (2012) Metabolic engineering of *Escherichia coli* for the production of 1-propanol. *Metab. Eng.*, **14**, 477–486.
- Kanehisa, M. et al. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
- Khersonsky, O., and Tawfik, D.S. (2010) Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.*, **79**, 471–505.
- Kim, J. et al. (2019) Metabolomic and transcriptomic analyses of *Escherichia coli* for efficient fermentation of L-fucose. *Mar. Drugs*, **17**, 82.
- Kim, J., and Copley, S.D. (2012) Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc. Natl. Acad. Sci. USA*, **109**, E2856–E2864.
- King, J.R. et al. (2017) Designing a new entry point into isoprenoid metabolism by exploiting fructose-6-phosphate aldolase side reactivity of *Escherichia coli*. *ACS Synth. Biol.*, **6**, 1416–1426.
- Ko, Y.-S. et al. (2020) Tools and strategies of systems metabolic engineering for the development of microbial cell factories for chemical production. *Chem. Soc. Rev.*, **49**, 4615–4636.
- Koch, M. et al. (2017) Molecular structures enumeration and virtual screening in the chemical space with RetroPath2.0. *J. Cheminf.*, **9**, 64.
- Lechner, A. et al. (2016) The need for integrated approaches in metabolic engineering. *Cold Spring Harb. Perspect. Biol.*, **8**, a023903.
- Liu, H. et al. (2013) Biosynthesis of ethylene glycol in *Escherichia coli*. *Appl. Microbiol. Biotechnol.*, **97**, 3409–3417.
- Major Group (2020) 28: Chemicals and Allied Products | Occupational Safety and Health Administration. United States Department of Labor, Washington, DC.
- Martin, V.J.J. et al. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nat. Biotechnol.*, **21**, 796–802.
- Mou, Z. et al. (2021) Machine learning-based prediction of enzyme substrate scope: application to bacterial nitrilases. *Proteins Struct.*, **89**, 336–347.
- Nielsen, J., and Keasling, J.D. (2016) Engineering cellular metabolism. *Cell*, **164**, 1185–1197.
- Notebaart, R.A. et al. (2014) Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. USA*, **111**, 11762–11767.
- Notebaart, R.A. et al. (2018) Underground metabolism: network-level perspective and biotechnological potential. *Curr. Opin. Biotechnol.*, **49**, 108–114.
- Oberhardt, M.A. et al. (2016) Systems-wide prediction of enzyme promiscuity reveals a new underground alternative route for pyridoxal 5'-phosphate production in *E. coli*. *PLoS Comput. Biol.*, **12**, e1004705.
- Okano, K. et al. (2018) De novo design of biosynthetic pathways for bacterial production of bulk chemicals and biofuels. *FEMS Microbiol. Lett.*, **365**, 138.
- Orth, J.D. et al. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.*, **7**, 535.
- Orth, J.D. et al. (2010) What is flux balance analysis? *Nat. Biotechnol.*, **28**, 245–248.
- Pereira, B. et al. (2016) Engineering a novel biosynthetic pathway in *Escherichia coli* for production of renewable ethylene glycol. *Biotechnol. Bioeng.*, **113**, 376–383.
- Pharkya, P. et al. (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.*, **14**, 2367–2376.
- Pickens, L.B. et al. (2011) Metabolic engineering for the production of natural products. *Annu. Rev. Chem. Biomol. Eng.*, **2**, 211–236.
- Pontrelli, S. et al. (2018) *Escherichia coli* as a host for metabolic engineering. *Metab. Eng.*, **50**, 16–46.
- Porokhin, V. et al. (2021) Analysis of metabolic network disruption in engineered microbial hosts due to enzyme promiscuity. *Metab. Eng. Commun.*, **12**, e00170.
- Python Software Foundation. *Python Language Reference, Version 3.6*. <http://www.python.org>.
- Robinson, S.L. et al. (2020) Machine learning-based prediction of activity and substrate specificity for OleA enzymes in the thiolase superfamily. *Synth. Biol.*, **5**, ysaa004.
- Rosenberg, J., and Commichau, F.M. (2019) Harnessing underground metabolism for pathway development. *Trends Biotechnol.*, **37**, 29–37.
- Salvy, P. et al. (2019) pyTFA and matTFA: a Python package and a Matlab toolbox for thermodynamics-based flux analysis. *Bioinformatics*, **35**, 167–169.
- Sánchez, B.J. et al. (2017) Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.*, **13**, 935.
- Subedi, K.P. et al. (2008) Role of GldA in dihydroxyacetone and methylglyoxal metabolism of *Escherichia coli* K12. *FEMS Microbiol. Lett.*, **279**, 180–187.
- Szappanos, B. et al. (2016) Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat. Commun.*, **7**, 11607.
- Tyzack, J.D. et al. (2019) Exploring chemical biosynthetic design space with transform-MinER. *ACS Synth. Biol.*, **8**, 2494–2506.
- Wang, H.H. et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894–898.
- Wang, L. et al. (2017) A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol.*, **2**, 243–252.
- Wang, T. et al. (2018) Genome-wide screening identifies promiscuous phosphatases impairing terpenoid biosynthesis in *Escherichia coli*. *Appl. Microbiol. Biotechnol.*, **102**, 9771–9780.
- Wannier, T.M. et al. (2020) Improved bacterial recombineering by parallelized protein discovery. *Proc. Natl. Acad. Sci. USA*, **117**, 13689–13698.
- Wery, T., and Petersen, G. (2004) *Top Value Added Chemicals from Biomass: Volume I – Results of Screening for Potential Candidates from Sugars and Synthesis*, Gas National Renewable Energy Lab., Golden, CO.
- Wu, Y. et al. (2016) Metabolic engineering strategies for co-utilization of carbon sources in microbes. *Bioengineering*, **3**, 10.
- Yim, H. et al. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol.*, **7**, 445–452.