

ARTICLE

In-depth characterization of genome-scale network reconstructions for the in vitro synthesis in cell-free systems

Lisa Katharina Schuh | Christian Weyler | Elmar Heinzle 

Biochemical Engineering Institute, Saarland University, Campus A1.5, Saarbrücken, Germany

Correspondence

Elmar Heinzle, Biochemical Engineering Institute, Saarland University, Campus A1.5, 66123 Saarbrücken, Germany.
Email: e.heinzle@mx.uni-saarland.de

Funding information

Bundesministerium für Bildung und Forschung Biotechnologie 2020+: Basistechnologien fuer eine naechste Generation biotechnologischer Verfahren, Grant/Award Number: FKZ 031P7238

Abstract

Cell-free systems containing multiple enzymes are becoming an increasingly interesting tool for one-pot syntheses of biochemical compounds. To extensively explore the enormous wealth of enzymes in the biological space, we present methods for assembling and curing data from databases to apply them for the prediction of pathway candidates for directed enzymatic synthesis. We use Kyoto Encyclopedia of Genes and Genomes to establish single organism models and a pan-organism model that is combining the available data from all organisms listed there. We introduce a filtering scheme to remove data that are not suitable, for example, generic metabolites and general reactions. In addition, a valid stoichiometry of reactions is required for acceptance. The networks created are analyzed by graph theoretical methods to identify a set of metabolites that are potentially reachable from a defined set of starting metabolites. Thus, metabolites not connected to such starting metabolites cannot be produced unless new starting metabolites or reactions are introduced. The network models also comprise stoichiometric and thermodynamic data that allow the definition of constraints to identify potential pathways. The resulting data can be directly applied using existing or future pathway finding tools.

KEYWORDS

biocatalysis, genome-scale network, metabolic pathway search, multienzyme systems, network curation, pan-organism network

1 | INTRODUCTION

The enzymatic potential of the numerous enzymes in nature is a most promising, extremely versatile, and powerful resource for creating powerful tools for the production of various interesting products. Besides the production in host organisms, synthesis using cell-free systems gains more and more interest. Particularly multistep biocatalysis seems only marginally explored today compared to its expected huge potential (Krauser & Weyler, 2013). Cell-free systems for the synthesis range from mixtures of isolated enzymes over multienzyme systems, for example, multienzyme complexes (S. Z. Wang et al., 2017) and enzyme cascades, to cell lysates (Endo & Koizumi, 2001) and permeabilized cells. In special cases, such systems are even combined with chemical synthesis in one pot (Groeger & Hummel, 2014).

The design of a multistep synthesis route does not only require the determination of the reaction sequence leading to the desired product, but also depends on numerous aspects such as substrate and cofactor supply or thermodynamics. For living cells, a recent review article discusses the state-of-the-art computational tools for design and reconstruction of metabolic pathways (Wang, Dash, Ng, & Maranas, 2017). To design such a pathway for cell-free biosynthesis is by far not developed to such a mature state. In particular, it seems almost impossible to explore manually all potentially feasible pathways and to determine which one is the most suitable for production.

The in silico path-finding and design methods all require a metabolic network model containing all required information from the host organisms of interest, such as enzyme, reaction, and thermodynamics data.

[Correction added on 14 November 2020, after first online publication: Projekt Deal funding statement has been added.]

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. *Biotechnology and Bioengineering* published by Wiley Periodicals LLC

There is an ever-growing plethora of biological databases with enzyme and reaction data of an ever-growing number of organisms that is suited for the reconstruction of genome-scale metabolic networks. One of the most popular databases is Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa & Goto, 2000; Kanehisa, Furumichi, Tanabe, Sato, & Morishima, 2016, 2018). However, despite the huge amount of data collected from primary literature that is carefully curated afterwards, the data is partly incomplete, and sometimes even inconsistent or erroneous. It is thus a challenge to handle these data and make them suitable for useful network reconstructions.

We already presented a computational tool to guide and support finding the most suitable synthesis path to a product (Blaß, Weyler, & Heinzle, 2017). We extended this study by developing a method of building network models from KEGG data, which is suitable for path-finding. We selected nine organism networks that are of interest primarily for their application in cell-free production. Some were selected because of peculiarities of the networks. Finally, a so-called pan-organism network was used lumping all metabolic reactions listed in KEGG in one single network.

2 | MATERIALS & METHODS

In the following, we give a short introduction to our path-finding method. We also present how to build network reconstruction models based on data found in biological databases, particularly KEGG.

2.1 | Path-finding

We already presented a method for finding candidates for suitable synthesis pathways in genome-scale metabolic network reconstructions starting from arbitrary substrates (Blaß et al., 2017). A *pathway* in our definition consists of two parts. First, the so-called *linear path* consists of a sequence of metabolites connected by reactions. It starts with a reaction that has one of the possible predefined start metabolites as a substrate and ends with a reaction that has the target metabolite T as a product. Second, there is the set of *supplying reactions*, which provide the substrates required by the reactions on the pathway that are not contained in the metabolite pool. All metabolites in this pool are considered freely available since they will be provided by the specified pathway reactions (see Section 2.2).

The path-finding algorithm is based on a mixed-integer linear program (MILP) and combines graph-based path-finding and reaction stoichiometry (Pey, Prada, Beasley, & Planes, 2011). The method is elaborated in detail in Blaß et al. (2017). Figure 1 shows an exemplary pathway illustrating a possible solution of the MILP. The pathway shown is a feasible synthesis pathway to the target T (depicted as red octagon). Metabolites in the figure are depicted as squares, where large squares represent metabolites in arcs (see

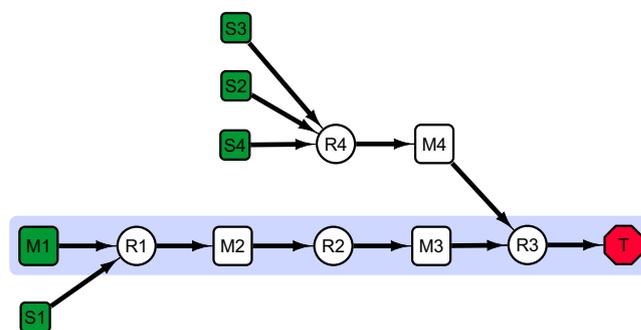


FIGURE 1 Exemplary pathway illustrating a feasible pathway to the target metabolite T (red octagon). Large squares: metabolites with arcs (see Section 2.2); small squares: cofactors/inorganic metabolites; green: metabolites from the metabolite pool (see Section 2.2); circles: reactions; blue background: linear path; R4: supplying reaction [Color figure can be viewed at wileyonlinelibrary.com]

Section 2.2) and small squares represent cofactors and inorganic metabolites. Reactions are represented by circles. The linear path of the pathway is marked with a blue background. Metabolites S1 to S4 and M1 (marked in green) are contained in the metabolite pool (see Section 2.2) and are thus initially available. As M4, which is required by reaction R3, is not available from the metabolite pool, R4 is needed as a supplying reaction producing it.

In addition to the 17 constraints of the MILP presented in Blaß et al. (2017), we added a constraint which prevents the use of a reaction in the pathway (more precisely, the supplying reactions) that consumes the target. This constraint is necessary to prevent cycles formed by a reaction belonging to the linear path that produces the target and a supplying reaction consuming the target to produce a precursor, which is consumed by a reaction on the linear path. It thus prevents pathways for which the target has to be already present in at least catalytic amounts. An example for such an undesired pathway is shown in Figure 2. In this example, the target T needs to be consumed by reaction R6 to form metabolite M4, which is required by reaction R5 to produce the target T.

The complete MILP is listed in Material S1.

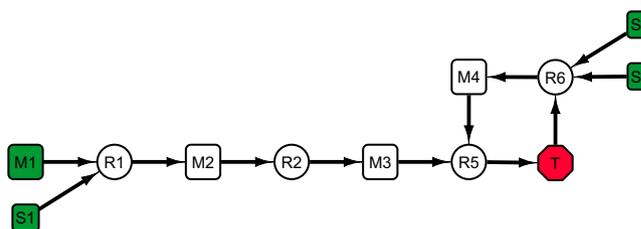


FIGURE 2 Exemplary pathway illustrating a pathway to the target metabolite T (red octagon) where T needs to be consumed to produce M4. Large squares: metabolites with arcs; small squares: cofactors/inorganic metabolites; green: metabolites from the metabolite pool; circles: reactions; blue background: linear path; R4 is a supplying reaction. This pathway example is not a valid synthesis pathway candidate for T [Color figure can be viewed at wileyonlinelibrary.com]

2.2 | Model building

In the following, we define the different parts of our network reconstruction and model based on KEGG data.

The reactions and metabolites in the model are given as lists of KEGG REACTION and COMPOUND ids (Kanehisa et al., 2016). The reactions and metabolites are connected by arcs, which are derived from reactions.

The metabolites in the model are categorized into sets that are treated differently in the path-finding algorithm. One set consists of potential *start metabolites*. These are all metabolites in the model that can be used as the start of the linear path of a pathway candidate. Metabolites in this category are automatically determined and have a molecular mass smaller than 300 and occur in arcs. The so-called *basis metabolites* are expert-curated metabolites, which are inexpensive, easily available, and are often hubs in the arc network, such as D-glucose (C00031) or pyruvate (C00022). The *cofactors* (e.g., ATP [C00002], NADH [C00004] and so forth) and *inorganics* such as water (C00001), oxygen (C00007) or CO₂ (C00011) are a set of expert-curated metabolites that are considered as freely available if they are required as substrates in reactions, but are not part of the reaction chain. They are thus excluded from the arcs to prevent biologically meaningless shortcuts in the pathways. All metabolite sets are disjoint, except for the basis metabolites that form a subset of the start metabolites. The *metabolite pool* is the superset of metabolites that are considered as freely available. It is made up of start metabolites, basis metabolites, cofactors, and inorganic metabolites. Further details on the different categories are given in Blaß et al. (2017).

For each reaction, there is a set of arcs, which are substrate–product pairs of a reaction. There are different strategies to derive the arcs from a reaction. The straightforward method is using all possible combinations (i.e., the cross product) of substrates and products of a reaction. It is however more useful to use meaningful substrate–product pairs, such as reactant pairs. A reactant pair is a substrate–product pair with both parts having atoms or atom groups in common that preserves the chemical substructures of the reactants through the reaction (Kotera et al., 2004; Kotera, Okuno, Hattori, Goto, & Kanehisa, 2004). The reactant pairs are defined in the KEGG RCLASS database, which classifies reactions based on the chemical structure patterns of their substrate–product pairs (Muto et al., 2013). Only those reactant pairs are used for the arcs that do not contain any metabolite from the cofactor and inorganics list. This means, however, that reactions involving metabolites from this list are still represented by the remaining arcs. A more detailed discussion on the arcs can be found in Section 1.4 of Material S2 (Tables B12–B16). The arc graph of the model is a directed graph $G = (V, E)$, where V is the set of metabolites and E is the set of arcs between these metabolites.

The model also contains a stoichiometric matrix, where each row corresponds to a metabolite in the model and each column indicates a reaction. An entry in the matrix is the stoichiometric coefficient of the metabolite in the respective reaction.

When using KEGG COMPOUND and KEGG REACTION data for a network reconstruction, some obstacles have to be addressed.

One of them is reaction directionality. For the reactions contained in KEGG, the reaction directions are not indicated in the database entries. There is thus a need for further reaction data to annotate directionality. To do so, we use the component contribution method of the biochemical thermodynamics calculator eQuilibrator (Flamholz, Noor, Bar-Even, & Milo, 2012; Noor, Haraldsdóttir, Milo, & Fleming, 2013) to compute the $\Delta_r G^m$ value (the change of the Gibbs free energy of a reaction at a given pH of 7 and ionic strength I in 1 mM concentration of the reactants) for each reaction in the network and infer if the respective reaction is reversible. Reactions with $|\Delta_r G| \leq 15$ kJ/mol are designated as reversible. In biological systems as well as in most biosynthetic setups, concentrations of substrates and products often differ by several orders of magnitude. This significantly influences reaction reversibility. As these effects cannot be adequately considered given the size of the networks presented in this study and the unknown kinetics, the $\Delta_r G$ value of 15 kJ/mol was chosen as a consensus value to determine reaction reversibility. This somewhat arbitrary value represents a compromise between the assumption of reversibility of all reactions and a more stringent restriction with a $\Delta_r G$ value of less than 15 kJ/mol that would potentially exclude feasible biosynthetic routes with concentrations of intermediates adjusting in a running system. The value was set after a series of simulations and expert inspection of results. However, the user of our tool can freely set the $\Delta_r G$ cutoff to meet the needs of his specific investigation. The reactions are added to the model in the respective direction(s), which means that for each reversible reaction, we get two reactions in the respective directions. Another obstacle is the inconsistent use of identifiers for metabolites. In some reaction equations, the KEGG COMPOUND (C) identifiers are used, and in others, the G identifiers from the KEGG GLYCAN structure database. As we do not consider glycans, those reactions are excluded. For polymerization reactions, the reaction stoichiometry in KEGG is not expressed in distinct numbers. Such reactions are not applicable for our method where the coefficients in the stoichiometric matrix are required to be integer numbers.

We did not generally exclude membrane-associated reactions. To our knowledge, it is not sufficiently clear whether and to what extent intracellular as well as extracellular membrane-associated enzymes are active in permeabilized cells. In earlier work, we could, however, experimentally show that megasynthases producing a circular oligopeptide can be kept active in permeabilized cells in contrast to cell extracts where activities could not be detected (Weyler & Heinzle, 2017). The exact reasons were not identified but could potentially be related to yet unknown membrane association. On the other hand, in selectively permeabilized eukaryotes, the organelles including membrane reactions remain intact and functional (e.g., Nicolae, Wahrheit, Nonnenmacher, Weyler, & Heinzle, 2015).

We thus have to filter the KEGG data before building a model. Figure 3 shows the filtering steps to obtain the reactions suitable for building a reconstruction of a pan-organism network encompassing reactions from all organisms and also for organism-specific networks.

The filtering starts with all 11,196 reactions in KEGG REACTION. First, the reactions with invalid reactants are removed, which are reactants that do not have a C identifier. The 10,764 remaining reactions

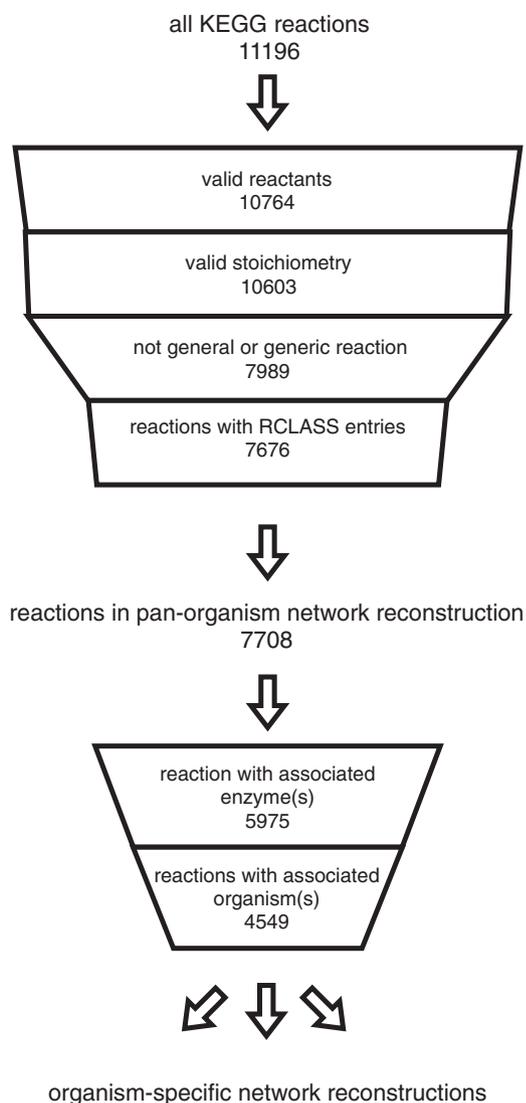


FIGURE 3 Reaction filtering from all reactions in Kyoto Encyclopedia of Genes and Genomes (KEGG) to the set of reactions for building the pan-organism network reconstruction and the organism-specific models. The reactions are filtered in the given order. The numbers indicate how many reactions stay after filtering. The width of the box bases are proportional to the number of reaction that remain after filtering

are further trimmed down to 10,603 reactions with valid stoichiometry, where all reactants have integer stoichiometric coefficients. From these, reactions that are generic or contain generic reactants (i.e., the database entry has a comment containing 'generic,' 'incomplete,' or 'general') are removed, sparing 7,989 reactions. After removing those without any reaction class annotations, 7,676 reactions remain in the pan-organism model, which corresponds to about 69% of all KEGG reactions.

To build the organism-specific models, the organism annotation for the genes of the enzymes catalyzing those reactions is used. From the 7,676 reactions in the pan-organism network reconstruction, KEGG has Enzyme Commission numbers associated with 5,975 reactions. Out of those, 4,549 (76%) have enzymes whose genes are annotated with organisms. These reactions are the basis of the organism-specific network model reconstructions.

Our network reconstruction workflow filters out ill-formed reaction entries in KEGG. However, we do not include a gap filling step. This would require large manual efforts that are not in the scope of this study.

Possible target metabolites in KEGG for the computation of synthesis pathway candidates are determined automatically. A target metabolite is a metabolite in the respective model that is not a dedicated start or basis metabolite. It also has to appear as a product in at least one arc in the network, so it could be potentially produced. We predict potentially producible targets in a given model by determining its feasible reactions, that is, reactions for which potentially all substrates are available or producible. The feasible reactions are obtained by initially starting with the set of metabolites consisting of the model's start metabolites and cofactors/inorganics. With these metabolites, all reactions that are feasible are determined by checking for each reaction that has not already been added to the set of feasible reactions if all substrates are available. The products of these feasible reactions are added to the set of metabolites. This step is repeated until no new substrates are added. The resulting set of reactions is then a subset of the model's reactions that potentially are feasible.

The next step is to do a reachability screening in the arc graph of the model. To do so, we add a node representing an artificial start metabolite that is connected to all potential start metabolites. From there, we do a breadth-first search in the graph. Breadth-first search is a suitable algorithm for exploring a graph. The search starts with a source vertex and discovers all neighboring vertices with the present depth before discovering the next depth-level vertices (Cormen, Leiserson, Rivest, & Stein, 2009). The potentially producible targets are those targets that are connected with the start node by a path (a sequence of edges that connect vertices) and that are produced by any of the feasible reactions.

2.3 | Computational details

The model data are based on KEGG release 90.1, May 1, 2019. The code for model building and statistics is written in Python 2.7, the code for the thermodynamics is written in Python 3.6 using the eQuilibrator API (Flamholz et al., 2012). We furthermore used the packages—graph-tool (Peixoto, 2014) and Matplotlib (Hunter, 2007). The path-finding tool was run on MATLAB R2019a with IBM CPLEX Studio 12.9. All computations were carried out on a 2.5 Ghz Intel Core i7 with 32 GB of RAM.

The software used in this study is available at <https://github.com/mecatsb>, where the repository `mecat` contains the path-finding tool and the repository `mecatpy` contains the code used for the pathway analysis as well as the organism models. Release v1.0 contains the code version used in this study.

3 | RESULTS

We first present the organisms and models used in our study and then discuss some interesting properties of these models. We finally present and discuss the results of our path-finding analysis.

3.1 | Models

For each organism in the KEGG Organisms database, we build an organism-specific network model, as described in Section 2.2. Figure 4 shows the number of reactions in KEGG that are annotated for the specific organism together with the number of reactions that are part of the organism-specific network reconstruction. The organisms are sorted in descending order with respect to the number of annotated reactions in the model. The order in which the reactions are filtered is following the procedure shown in Figure 3.

On average, 67% of the reactions in KEGG that are annotated with an organism end up in an organism-specific model (see insert in Figure 4). The reason for this is the filtering of all reactions according to the filter constraints shown in Figure 3 and discussed in detail in Section 2.2. Figure 3 shows that the majority of the discarded reactions are general and/or generic or contain generic reactants.

In addition to the pan-organism network model, we chose nine organism-specific models for all network and pathway analyses as examples. Table 1 lists the organisms, which were chosen primarily for their importance in biotechnological production as well as in scientific research. CHO, the permanent cells of the ovary of a Chinese hamster *C. griseus* were originally isolated already in 1957. CHO is serving as a model cell line for metabolic studies. Most importantly, however, it is most frequently used for the industrial heterologous production of therapeutic proteins (Lalonde & Durocher, 2017; Wurm, 2004). The application of animal cells for biosynthetic purposes is easier starting from cell lines like CHO rather than cells from primary tissues. *Escherichia coli* is probably the most important model organism and is used in all kinds of areas spanning from basic molecular biological work to industrial applications (Pontrelli et al., 2018). *Vibrio natriegens* is an extremely fast growing marine bacterium that recently got increasing interest. Due to its duplication time of 10 minutes, it has been in the focus of molecular biology research, for example, for protein production

also in cell-free systems (Failmezger, Scholz, Blombach, & Siemann-Herzberg, 2018; Hoffart et al., 2017). *Pseudomonas putida* is known for its diverse biodegradation and biosynthetic capabilities (Loeschcke & Thies, 2015; Nickel, Chavarria, Danchin, & de Lorenzo, 2016; Poblete-Castro, Becker, Dohnt, DosSantos, & Wittmann, 2012). *Myxococcus xanthus* is a model organism for studying social behavior of bacteria with extended signaling networks and secondary metabolite production (Wrótniak-Drzewiecka, Brzezińska, Dahm, Ingle, & Rai, 2016). *Saccharomyces cerevisiae* is probably the most important eukaryotic model micro-organism used very widely and already for a long time for the production of ethanol in alcoholic beverages and biofuel. It is also widely discussed for the production of other metabolites and its broad application is supported by a large toolbox for metabolic engineering (Krivoruchko & Nielsen, 2015; Nielsen, 2019; Steensels et al., 2014). The yeast *Schizosaccharomyces pombe* is a model organism primarily used in molecular and cell biology but is recently also discussed as promising candidate for the expression and secretion of heterologous proteins (Takegawa et al., 2009). *Corynebacterium glutamicum* is a most important micro-organism in the industrial scale production of amino acids but also other metabolic products (Becker, Gießelmann, Hoffmann, & Wittmann, 2016). While these organisms have been used in a vast range of production processes, they are also well understood and we assume that KEGG data on these organisms is relatively complete and accurate. *Mycoplasma penetrans* has the smallest genome of known organisms and its metabolism is very limited (Sasaki et al., 2002). From the present view, the most important organisms for cell-free synthesis are *E. coli*, *S. cerevisiae*, *P. putida*, and *M. xanthus*. We exclude plants and algae from our species models since they seem less applicable from the present view on cell-free biocatalysis.

All model data are part of the GitHub repository <https://github.com/mecatpb/mecatpb>.

Table 2 shows the number of potential targets for the respective model as defined in Section 2.2. The arc reachable targets are those targets that are connected to a basis metabolite via an arc path, which is determined by breadth-first search. The feasible targets are targets that are products of feasible reactions as described in Section 2.2. The set of potentially producible targets is the intersection of the targets that are connected to a basis metabolite via an arc path and the targets that are products of the feasible reactions.

Table 2 shows that a large portion of potential targets is not connected to any of the basis metabolites in the model. For all models, about 32% (in the *S. pombe* model spo) to 43% (in the pan-organism model kegg) of all potential targets are potentially producible targets. This means that for all other potential targets, a synthesis pathway cannot be found, as a path is a required part of a valid solution. We will elaborate the reasons for this drastic reduction later in this study.

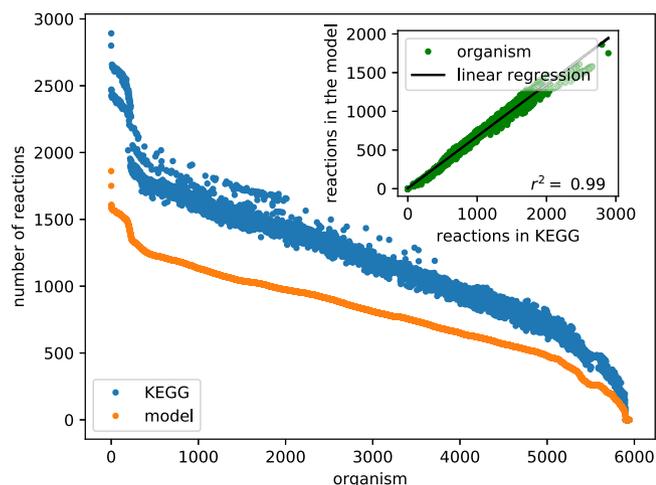


FIGURE 4 Comparison of the total number of reactions and the number of reactions selected for the models for all organisms annotated in KEGG. KEGG, Kyoto Encyclopedia of Genes and Genomes [Color figure can be viewed at wileyonlinelibrary.com]

3.2 | Network model analysis

We first present some basic properties of the arc graphs of the different organism models.

TABLE 1 Models for the studies

Model	Name	Reactions (KEGG/model/ reversible)	Feasible reactions	Metabolites (model/basis)
kegg	Pan-organism network model	11,196/7,676/2,934	5,467	6,473/39
cge	<i>Cricetulus griseus</i> (Chinese hamster)	2,616/1,555/639	922	1,485/39
eco	<i>Escherichia coli</i> K-12 MG1655	1,775/1,225/511	950	1,191/39
vna	<i>Vibrio natriegens</i>	1,690/1,193/489	888	1,180/39
ppun	<i>Pseudomonas putida</i> NBRC 14164	1,683/1,199/469	766	1,240/37
mxs	<i>Myxococcus xanthus</i>	1,492/1,021/428	684	1,077/36
sce	<i>Saccharomyces cerevisiae</i> (budding yeast)	1,543/1,020/403	655	1,031/39
spo	<i>Schizosaccharomyces pombe</i> (fission yeast)	1,408/905/378	592	915/39
cgb	<i>Corynebacterium glutamicum</i> ATCC 13032 (Bielefeld)	1,122/792/318	534	845/38
mpe	<i>Mycoplasma penetrans</i>	371/236/101	166	314/22

Note: The model names are derived from the KEGG organism codes, except for the pan-organism network model, which is named kegg. The number of reactions in KEGG refers to the number of reactions that are annotated for the respective organism. The number of reversible reactions is the corresponding subset of the reactions in the model. The feasible reactions are determined as described in Section 2.2 based on the set of basis metabolites as start metabolites. The basis metabolites are selected as described in Section 2.2.

Abbreviation: KEGG, Kyoto Encyclopedia of Genes and Genomes.

Figure B1 in Material S2 shows the node degree distributions of the arc graphs of the different organism network reconstructions. The *degree* of a node is the number of edges leaving it (out-degree) plus the number of edges entering it (in-degree). Tables B1–B10 in Material S2 list the hubs with the top five occurrences of each network. As expected, pyruvate, L-glutamate, D-glyceradehyde 3-phosphate, and acetyl-CoA are in almost all cases metabolites with highest node degrees. *M. penetrans* (mpe), having the smallest network of all studied here, differs most significantly from all others both in the types of metabolites with highest node degrees as well as in the generally small numbers of node degrees (<13). In the pan-organism network model (kegg), trans,trans-farnesyl diphosphate has an exceptionally high node degree (107), that is, however, mostly originating from plant metabolism. In kegg,

pyruvate is by far the most connected metabolite with a node degree of 167. The outstanding role of only a few metabolites is most strikingly seen in Figure B1 of Material S2.

The sizes of the arc graphs together with the average node degrees, standard deviation of the distribution are listed in Table B11 of Material S2. It is interesting to see that the average node degrees vary only from 2.37 to 3.14 for individual organisms and 3.3 for kegg, the pan-organism network.

Table 3 lists the number of connected components of the arc graphs in the respective models and the size of the largest connected component, respectively. A connected component in the graph is a subgraph where each vertex in the subgraph is connected to each other vertex in the subgraph by a path (Cormen et al., 2009).

TABLE 2 Number of potential targets for each organism model based on basis metabolites as possible start metabolites

Model	Potential targets	Arc reachable targets	Feasible targets	Potentially producible targets	% of potential targets
kegg	5,441	3,017	2,412	2,325	43
cge	1,128	437	358	333	30
eco	878	419	376	351	40
vna	865	397	348	328	38
ppun	902	380	317	293	32
mxs	777	320	281	266	34
sce	713	268	243	227	32
spo	637	264	216	201	32
cgb	598	261	215	200	33
mpe	184	70	69	56	30

Note: Arc reachable targets: targets that are connected to a basis metabolite via an arc path; feasible targets: targets that are products of feasible reactions as described in Section 2.2; potentially producible targets: targets that are connected to a basis metabolite via an arc path and that are products of feasible reactions and are thus realistic targets, intersection of the former two columns of the table; % of potential targets: percentage of potentially producible targets in relation to the total number of potential targets.

TABLE 3 Number of components in the models with the number of metabolites in the largest component

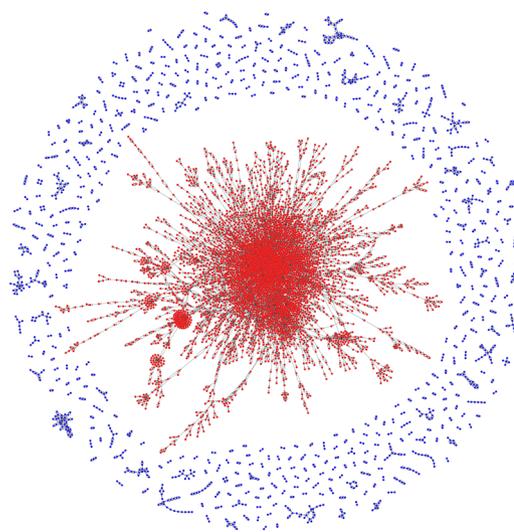
Model	Number of components	Size of largest component	Components with basis metabolites	Metabolites in components with basis metabolites
kegg	481	4,612	1	4,612 (74%)
cge	186	754	4	763 (55%)
eco	139	766	2	768 (69%)
vna	146	726	1	726 (66%)
ppun	157	761	2	763 (66%)
mxa	159	587	3	595 (60%)
sce	182	478	1	478 (50%)
spo	162	438	1	438 (52%)
cgb	115	457	2	459 (60%)
mpe	57	76	7	141 (54%)

Note: The fourth column lists the number of components containing basis metabolites. The last column shows the number of metabolites that belong to a component containing basis metabolites. The percentage of those metabolites in relation to the number of metabolites in the arc graph is shown in parentheses.

The smallest connected components contain two vertices in all models. This is by definition the smallest component size as the arc graph does not contain metabolites without any arcs. We furthermore list the number of components containing basis metabolites; as well as the total number of metabolites in all those components with the percentage of those metabolites in relation to the number of metabolites in the arc graph (in parentheses). These numbers give information on how much of each network is possibly reachable from the designated start points, since a potentially producible target has to be connected to any of the predefined basis metabolites via an arc path. Table 3 shows that between half and two-third of the metabolites in a model's arc graph are contained in a component with basis metabolites.

Exemplarily, Figure 5 shows the arc graph of the pan-organism model kegg. The arc graph consists of a large main component and a large number of small components. Components in red are components containing potential start metabolites, whereas the components in blue are the so-called satellite components without start metabolites. The arc graphs of the other models are shown in Figure B2 of Material S2. Figure B3 of Material S2 shows the arc graph component histograms.

There are several reasons for isolated components in a model. The first reason is missing annotation in the data on which the model is based. This could be improved by using manually compiled and curated network reconstructions with gap filling. Several reactions in KEGG are formulated as general reactions and/or are reactions containing generic compounds. Some of the often numerous reactions summarized in such reactions are explicitly listed in KEGG. An even larger number could, in principle, be added, for example, from BRENDA (Jeske, Placzek, Schomburg, Chang, & Schomburg, 2018). Some reactions involve

**FIGURE 5** Arc graph of the pan-organism model kegg. Red: components containing potential start metabolite; blue: satellite components without start metabolites [Color figure can be viewed at wileyonlinelibrary.com]

additional proteins that transfer electrons or groups or use covalently bound cofactors as, for example, nicotinamide adenine dinucleotide phosphate. These are filtered out in the model building process. As we do not include such reactions in our model, some metabolic pathways could be cut off. Another reason is that a component is really isolated.

In Material S3, we list all components identified in the kegg model. The 6,246 metabolites are grouped in 481 components. The largest component connected to start metabolites comprises 4,612 metabolites and is represented in the center of Figure 5. All other components with a size of 5 or more metabolites were investigated in more detail (Material S4). They comprise 682 metabolites in 69 components. We could identify some typical families related to biochemical characteristics (Material 5). Reactions of xenobiotic compounds, for example, drugs, were most prominent with 14 components with 141 metabolites followed by polyketides (10/109), carbohydrate-derived metabolites (10/86), terpenoids (9/112), compounds with gonane tape nucleus (7/74), fatty acid and lipids related compounds (7/59), and flavonoids (4/54). Xenobiotics are inherently not listed in the starting metabolites. Some of these families have often general reactions or involve generic metabolites, for example, metabolites contain a group -R that is not explicitly specified. R is later cleaved off the metabolite. Smaller components (<5) were not analyzed in detail but could often serve as missing links in larger pathways once the connecting reactions could be defined following the criteria specified in Section 2.2.

3.3 | Reachability analysis

We determined the target reachability in the organism-specific networks by testing the existence of a pathway candidate to each

possible target starting with basis metabolites using our MILP presented in Section 2.1. Figure 6 shows for each model the percentage of targets for which a synthesis pathway candidate has been identified and for which a pathway candidate is not accessible and why. The raw data for the figure is listed in Table B17 of Material S2. In the following, we discuss the different fractions in more detail.

The blue and orange fractions represent the targets for which a synthesis pathway candidate has been identified in the respective models. The targets represented by the orange fractions have been predicted to have a pathway candidate. This means that they can be produced by feasible reactions of the model and they are connected to at least one of the predefined basis metabolites by a path in the arc graph (see Section 2.2). An example for this category is UDP-glucose (Material S6, Section F.1). However, the targets represented by the blue fractions have not been predicted to be feasible, despite having a synthesis pathway candidate. For those targets, we found that most of the pathway candidates calculated with the MILP include a direct cycle formed by supplying reactions that use metabolites that are not in the metabolite pool. In a mathematical sense, it is valid to consume a metabolite as long as its overall balance is zero. However, in real world applications, this would not be correct since the metabolite has to be

present in at least catalytic amounts already at the start of the reaction. An example for such a pathway is the pathway candidate for the 5-methyl-5,6,7,8-tetrahydromethanopterin (C04488) production in the pan-organism network model kegg (Material S6, Section F.2). The pathway requires coenzyme F420 (C00876) and reduced coenzyme F420 (C01080), which are neither metabolites nor cofactors, and thus are not part of the metabolite pool. They thus have to be produced by the reactions of the pathway.

The green, red, purple, and brown fractions represent targets without any pathway candidate. In the following, we will discuss the different reasons for this.

With the help of breadth-first search, we found that the targets represented by the green and red fractions are not connected to any of the potential start metabolites via an arc path. Therefore, these targets cannot have a pathway candidate, since a path from a start metabolite to the target is mandatory, as stated in Section 2.1.

The targets belonging to the green category are not part of a component containing potential start metabolites. In our pan-organism model kegg, this is the case for proansamycin X. Component 134 in Material S4 shows that there is no reaction in KEGG producing proansamycin X (C12176) from 3-amino-5-hydroxybenzoate (C12107), which belongs to a component with start metabolites (Material S3). The situation could be improved by using manually compiled and curated network reconstructions with gap filling, for example, for metabolites of the earlier discussed polyketide, flavone, and terpenoid families (see also Material S5). As outlined in Section 2.2, we only did some minor generic curation which has the purpose of extracting meaningful data and removing ill-specified data. A comprehensive network reconstruction for an organism would require a lot of manual work encompassing more data sources including primary literature, which was not in the scope of this study. However, when using our path-finding method, the user can choose any network model that contains the information needed for path-finding, regardless of data origin.

The targets represented by the red fractions are contained in components with start metabolites but do not have a necessary arc path from a start metabolite to the target, such as riboflavin (C00255).

The targets represented by the purple fractions are connected to a potential start metabolite in the network via an arc path. However, this is not sufficient for a valid pathway candidate. In addition, the arcs have to be associated with reactions for which all substrates are available or producible to ensure that the pathway candidate is feasible (Blaß et al., 2017). However, for these targets, there is no reaction in the set of feasible reactions (see Section 2.2) that produces that target for the last one arc of the arc path, which means that the overall pathway is not feasible. Note that the other arc-reaction associations thus do not matter in this case. An example for such a target is biotin (C00120; Material S6, Section F.3).

The targets represented by the brown fractions are targets that are predicted to have pathway candidates as they are connected to predefined start metabolites by an arc path and are produced by feasible reactions. However, our path-finding algorithm could not determine valid pathway candidates. To explore the reasons for this, we list the feasible reactions of the respective models that produce these

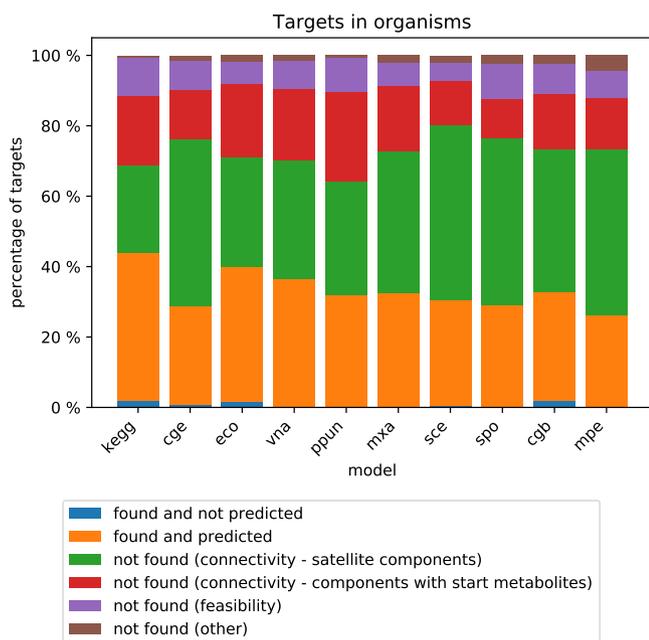


FIGURE 6 Analysis of the target search in the different organism models. blue: targets for which a pathway candidate has been found by our method, but that have not been predicted as feasible; orange: targets for which a pathway candidate has been found by our method; green: targets for which a candidate has not been found due to the absence of an arc path from any start metabolite to the target because the target is in a satellite component without start metabolite; red: targets for which a candidate has not been found due to the absence of an arc path from any start metabolite to the target (and the target is in a component with start metabolites); purple: targets for which a candidate has not been found due to the lack of a feasible reaction that produces the target; brown: targets for which a candidate has not been found due to other reasons that are discussed in the text [Color figure can be viewed at wileyonlinelibrary.com]

targets for each of those targets. For each of these reactions, we determine why it is not part of a pathway candidate. We identified three non-disjoint categories in which we can sort these reactions. To the first category, belong reactions that produce the target but do not have arcs containing the target. As discussed in Section 2.1, a valid pathway candidate has to include a reaction with an arc to the target. Reactions that produce the targets only from substrates that are designated cofactors or inorganic metabolites are also sorted to this category, as they are correctly predicted to be feasible. However, our path-finding method does not handle such pathways since a valid pathway candidate requires at least one arc by definition and there are no arcs containing cofactors and inorganic metabolites. The second category encompasses reactions that do have an arc to the target, but require a substrate that is also a target for which no pathway candidate has been identified with our method. The reactions in the third category cannot be used in a pathway candidate because of a constraint in the MILP, which excludes pathways that use reactions consuming the target, as discussed in Section 2.1. There is no valid sequence of reactions with arcs that is feasible without using supplying reactions that consume the target. An example for a target of the brown fraction is 5'-methylthioadenosine (C00170), where the reactions producing this target belong to the first two categories discussed above (Material S6, Section F.4).

To illustrate the different target categories, the example arc graph in Figure 7a depicts examples for each of the categories. Note that, for the sake of clarity, the depicted arc graph has additional vertices for the cofactors (small circles), which would normally not be part of the graph. The potential start metabolites A and B are depicted by hexagons, the potential targets E, F, G, and H by octagons. Figure 7b lists the reaction equations and the arcs belonging to these reactions. A valid pathway candidate to E consists of the reactions R1 and R3 (arcs 1, 2, and 4), since all needed substrates, that is, A, B, and Y, are available. E would thus be a target represented by the orange fractions in Figure 6.

For target F, it is not possible to find a pathway candidate because F is part of a graph component that does not include potential start metabolites. F is thus an example of the green fractions. Target G is part of the component which also includes the potential start metabolites. However, there is no arc path connecting metabolites A or B to G, which makes G a target represented by the red fractions.

Target H is an example for the purple fraction. To reach H, there are valid arc paths (e.g. $1 \rightarrow 5 \rightarrow 8$ or $2 \rightarrow 5 \rightarrow 8$), however, the last reaction belonging to arc 8 requires W as a substrate, which is not available.

4 | CONCLUDING REMARKS

Our presented method allows creating and characterizing genome-scale metabolic network reconstructions for the planning of biosynthetic production pathways using cell-free systems. The data are taken from biological databases, for example, KEGG. We also discussed typical problems in the context of network reconstruction and how these can be solved to obtain applicable network models. We used the presented method for establishing models for the network reconstruction of a pan-organism from the whole KEGG database as well as for several

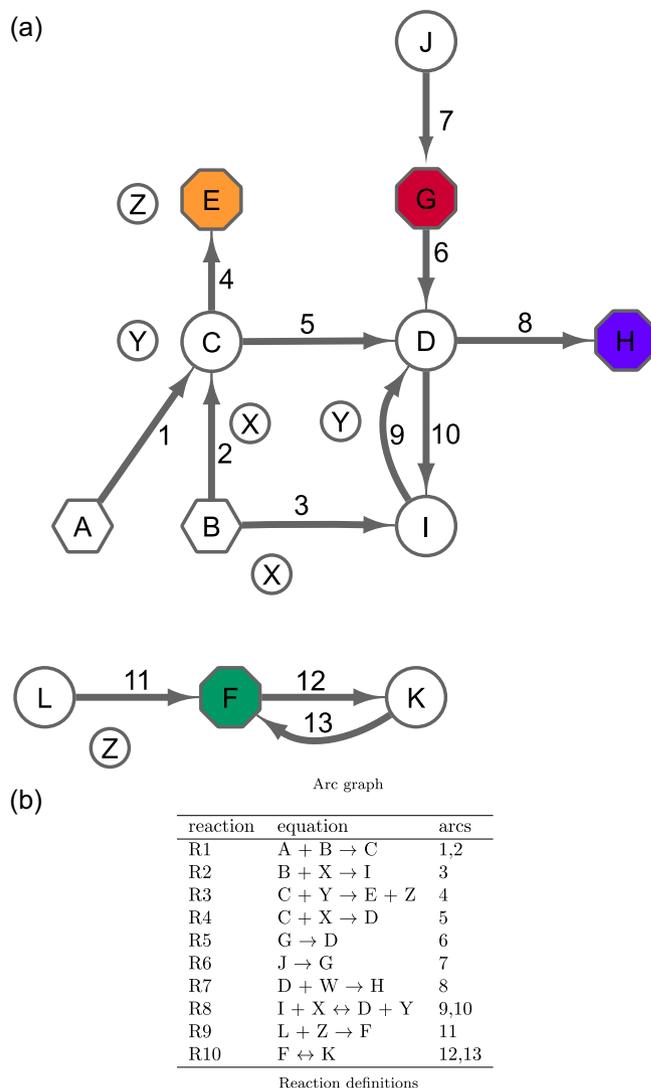


FIGURE 7 Arc graph with examples for the different target categories in Figure 6 and the corresponding reactions. (a) Small circles: cofactor metabolites; hexagons: potential start metabolites; octagons: potential targets; large circles: metabolites not in any of the previous categories. Orange: target of the orange category; green: target of the green category; red: target of the red category; purple: target of the purple category. The numbers on the arcs refer to the column 'arcs' in (b). (b) Reactions in the example network with their respective reaction equations [Color figure can be viewed at wileyonlinelibrary.com]

interesting model organisms. We also used our path-finding method based on a global optimization problem to compute pathway candidates for all possible target molecules in the models and demonstrated that our method yields correct and meaningful results and that it is widely applicable for all kinds of networks and network sizes. The increasing availability of larger-scale metabolic networks that are increasingly well curated, as is for example, already the case for *E. coli* and *S. cerevisiae* (Orth et al., 2011; Zomorodi & Maranas, 2010), will also increase the power of our method. Our network analysis method for multienzyme systems that do not have any cellular compartments particularly lacking a cell membrane differs significantly from published methods for whole cells with a defined link to the extracellular environment via

transport systems (S. Z. Wang et al., 2017; von Kamp, Thiele, Hädicke, & Klamt, 2017) or with models of microbial communities, for example, Magnúsdóttir and Thiele (2018).

The tools we presented are directly applicable to designing the synthesis of target compounds in cell-free systems. Our analysis tools—especially the feasibility prediction we described in Section 2.2—are useful tools to predict if a target could potentially be produced in a given model and could thus be used to quickly screen if a host organism or strain is potentially capable of producing a certain product directly.

If this is not the case, a comparison of biosynthesis pathways in a selected host organism and in the pan-organism is useful for identifying genetic engineering targets to create a production organism eventually. Our tools help to identify heterologous enzymes that might be candidates for insertion in the host organism chosen using genetic engineering to complete a desired pathway in that organism. Our tools also help to answer which substrates are required for a certain synthesis pathway.

Biosynthesis pathway candidates including stoichiometric and thermodynamic constraints can be determined with our presented path-finding algorithm presented earlier (Blaß et al., 2017). As reviewed in a recent publication (Lin, Warden-Rothman, & Voigt, 2019), various methods have already been published and are in development that additionally allow the identification of new reactions considering the promiscuity of many enzymes but also the chemical similarity of substrates of these enzymes.

Our network reconstructions are the basis for the identification of gaps in the network that would prohibit synthesis of a desired target. With our tools, it is possible to identify potential gap fillers from the pan-organism network, which can then be implemented in an organism of interest using genetic engineering. It is also possible to do manual directed gap filling in the pan-organism network, for example, by considering generic reactions, reactions not contained in KEGG, or expert reasoning.

Overall, our tools and networks are a suitable basis for focused and directed experimental work and the implementation of the synthesis of target compounds in cell-free systems.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interests.

ACKNOWLEDGMENT

This study was supported by the Federal Ministry of Education and Research, Germany (BMBF) within the initiative “Biotechnologie 2020+: Basistechnologien für eine nächste Generation biotechnologischer Verfahren” (FKZ 031P7238). Open access funding enabled and organized by Projekt DEAL.

ORCID

Elmar Heinzle  <http://orcid.org/0000-0002-1689-9300>

REFERENCES

- Becker, J., Gießelmann, G., Hoffmann, S. L., & Wittmann, C. (2016). *Corynebacterium glutamicum* for sustainable bioproduction: From metabolic physiology to systems metabolic engineering, *Synthetic Biology—Metabolic Engineering* (pp. 217–263). Cham, Switzerland: Springer. https://doi.org/10.1007/10_2016_21
- Blaß, L. K., Weyler, C., & Heinzle, E. (2017). Network design and analysis for multi-enzyme biocatalysis. *BMC Bioinformatics*, *18*(1), 366. <https://doi.org/10.1186/s12859-017-1773-y>
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. Cambridge: MIT Press.
- Endo, T., & Koizumi, S. (2001). Microbial conversion with cofactor regeneration using genetically engineered bacteria. *Advanced Synthesis & Catalysis*, *343*(6-7), 521–526. [https://doi.org/10.1002/1615-4169\(200108\)343:6/7<521::AID-ADSC521>3.0.CO;2-5](https://doi.org/10.1002/1615-4169(200108)343:6/7<521::AID-ADSC521>3.0.CO;2-5)
- Failmezger, J., Scholz, S., Blombach, B., & Siemann-Herzberg, M. (2018). Cell-free protein synthesis from fast-growing *Vibrio Natriegens*. *Frontiers in Microbiology*, *9*, 1146. <https://doi.org/10.3389/fmicb.2018.01146>
- Flamholz, A., Noor, E., Bar-Even, A., & Milo, R. (2012). eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Research*, *40*(D1), D770–D775. <https://doi.org/10.1093/nar/gkr874>
- Groeger, H., & Hummel, W. (2014). Combining the ‘two words’ of chemocatalysis and biocatalysis towards multi-step one-pot processes in aqueous media. *Current Opinion in Chemical Biology*, *19*, 171–179. <https://doi.org/10.1016/j.cbpa.2014.03.002>
- Hoffart, E., Grenz, S., Lange, J., Nitschel, R., Müller, F., Schwentner, A., & Blombach, B. (2017). High substrate uptake rates empower *Vibrio natriegens* as production host for industrial biotechnology. *Applied Environmental Microbiology*, *83*(22), <https://doi.org/10.1128/AEM.01614-17>
- Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jeske, L., Placzek, S., Schomburg, I., Chang, A., & Schomburg, D. (2018). Brenda in 2019: A european elixir core data resource. *Nucleic Acids Research*, *47*(D1), D542–D549. <https://doi.org/10.1093/nar/gky1048>
- 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>
- Kanehisa, M., & Goto, S. (2000). Kegg: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2016). Kegg: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, *45*(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2018). New approach for understanding genome variations in kegg. *Nucleic Acids Research*, *47*(D1), D590–D595. <https://doi.org/10.1093/nar/gky962>
- Kotera, M., Okuno, Y., Hattori, M., Goto, S., & Kanehisa, M. (2004). Computational assignment of the ec numbers for genomic-scale analysis of enzymatic reactions. *Journal of the American Chemical Society*, *126*(50), 16487–16498. <https://doi.org/10.1021/ja0466457>
- Kotera, M., Hattori, M., Oh, M. A., Yamamoto, R., Komeno, T., Yabuzaki, J., & Kanehisa, M. (2004). Rpair: A reactant-pair database representing chemical changes in enzymatic reactions. *Genome Informatics*, *15*, P062.
- Krauser, S., Weyler, C., Blaß, L. K., & Heinzle, E. (2013). Directed multistep biocatalysis using tailored permeabilized cells, *Fundamentals and Application of New Bioproduction Systems* (pp. 185–234). Berlin, Heidelberg, Germany: Springer. https://doi.org/10.1007/10_2013_240
- Krivoruchko, A., & Nielsen, J. (2015). Production of natural products through metabolic engineering of *Saccharomyces cerevisiae*. *Current*

- Opinion in *Biotechnology*, 35, 7–15. <https://doi.org/10.1016/j.copbio.2014.12.004>
- Lalonde, M. E., & Durocher, Y. (2017). Therapeutic glycoprotein production in mammalian cells. *Journal of Biotechnology*, 251, 128–140. <https://doi.org/10.1016/j.jbiotec.2017.04.028>
- Lin, G. M., Warden-Rothman, R., & Voigt, C. A. (2019). Retrosynthetic design of metabolic pathways to chemicals not found in nature. *Current Opinion in Systems Biology*, <https://doi.org/10.1016/j.coisb.2019.04.004>
- Loeschcke, A., & Thies, S. (2015). *Pseudomonas putida*—a versatile host for the production of natural products. *Applied Microbiology and Biotechnology*, 99(15), 6197–6214. <https://doi.org/10.1007/s00253-015-6745-4>
- Magnúsdóttir, S., & Thiele, I. (2018). Modeling metabolism of the human gut microbiome. *Current Opinion in Biotechnology*, 51, 90–96. <https://doi.org/10.1016/j.copbio.2017.12.005>
- Muto, A., Kotera, M., Tokimatsu, T., Nakagawa, Z., Goto, S., & Kanehisa, M. (2013). Modular architecture of metabolic pathways revealed by conserved sequences of reactions. *Journal of Chemical Information and Modeling*, 53(3), 613–622. <https://doi.org/10.1021/ci3005379>
- Nicolae, A., Wahrheit, J., Nonnenmacher, Y., Weyler, C., & Heinze, E. (2015). Identification of active elementary flux modes in mitochondria using selectively permeabilized cho cells. *Metabolic Engineering*, 32, 95–105.
- Nielsen, J. (2019). Yeast systems biology: Model organism and cell factory. *Biotechnology Journal*, 14(9), 1800421. <https://doi.org/10.1002/biot.201800421>. <https://onlinelibrary.wiley.com/doi/abs/10.1002/biot.201800421>
- Nikel, P. I., Chavarria, M., Danchin, A., & de Lorenzo, V. (2016). From dirt to industrial applications: *Pseudomonas putida* as a synthetic biology chassis for hosting harsh biochemical reactions. *Current Opinion in Chemical Biology*, 34, 20–29. <https://doi.org/10.1016/j.cbpa.2016.05.011>
- Noor, E., Haraldsdóttir, H. S., Milo, R., & Fleming, R. M. (2013). Consistent estimation of gibbs energy using component contributions. *PLoS Computational Biology*, 9(7), <https://doi.org/10.1371/journal.pcbi.1003098>
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology*, 7(1), <https://doi.org/10.1038/msb.2011.65>
- Peixoto, T. P. (2014). The graph-tool python library. *Figshare*, http://figshare.com/articles/graph_tool/1164194
- Pey, J., Prada, J., Beasley, J., & Planes, F. (2011). Path finding methods accounting for stoichiometry in metabolic networks. *Genome Biology*, 12(5), R49. <https://doi.org/10.1186/gb-2011-12-5-r49>
- Poblete-Castro, I., Becker, J., Dohnt, K., DosSantos, V. M., & Wittmann, C. (2012). Industrial biotechnology of *Pseudomonas putida* and related species. *Applied Microbiology and Biotechnology*, 93(6), 2279–2290. <https://doi.org/10.1007/s00253-012-3928-0>
- Pontrelli, S., Chiu, T. Y., Lan, E. I., Chen, F. Y. H., Chang, P., & Liao, J. C. (2018). *Escherichia coli* as a host for metabolic engineering. *Metabolic Engineering*, 50, 16–46. <https://doi.org/10.1016/j.ymben.2018.04.008>
- Sasaki, Y., Ishikawa, J., Yamashita, A., Oshima, K., Kenri, T., Furuya, K., & Hattori, M. (2002). The complete genomic sequence of *Mycoplasma penetrans*, an intracellular bacterial pathogen in humans. *Nucleic Acids Research*, 30(23), 5293–5300. <https://doi.org/10.1093/nar/gkf667>
- Steensels, J., Snoek, T., Meersman, E., Nicolino, M. P., Voordeckers, K., & Verstrepen, K. J. (2014). Improving industrial yeast strains: Exploiting natural and artificial diversity. *FEMS Microbiology Reviews*, 38(5), 947–995. <https://doi.org/10.1111/1574-6976.12073>
- Takegawa, K., Tohda, H., Sasaki, M., Idiris, A., Ohashi, T., Mukaiyama, H., & Kumagai, H. (2009). Production of heterologous proteins using the fission-yeast (*Schizosaccharomyces pombe*) expression system. *Biotechnology and Applied Biochemistry*, 53(4), 227–235. <https://doi.org/10.1042/BA20090048>
- Wang, L., Dash, S., Ng, C. Y., & Maranas, C. D. (2017). A review of computational tools for design and reconstruction of metabolic pathways. *Synthetic and Systems Biotechnology*, 2(4), 243–252. <https://doi.org/10.1016/j.synbio.2017.11.002>
- Wang, S. Z., Zhang, Y. H., Ren, H., Wang, Y. L., Jiang, W., & Fang, B. S. (2017). Strategies and perspectives of assembling multi-enzyme systems. *Critical Reviews in Biotechnology*, 37(8), 1024–1037. <https://doi.org/10.1080/07388551.2017.1303803>
- Weyler, C., & Heinze, E. (2017). Synthesis of natural variants and synthetic derivatives of the cyclic nonribosomal peptide luminide in permeabilized *E. coli* nisse and product formation kinetics. *Applied Microbiology and Biotechnology*, 101(1), 131–138.
- Wrótniak-Drzewiecka, W., Brzezińska, A. J., Dahm, H., Ingle, A. P., & Rai, M. (2016). Current trends in myxobacteria research. *Annals of Microbiology*, 66(1), 17–33. <https://doi.org/10.1007/s13213-015-1104-3>
- Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, 22(11), 1393. <https://doi.org/10.1038/nbt1026>
- Zomorodi, A. R., & Maranas, C. D. (2010). Improving the imm904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Systems Biology*, 4, 178. <https://doi.org/10.1186/1752-0509-4-178>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Schuh LK, Weyler C, Heinze E. In-depth characterization of genome-scale network reconstructions for the in vitro synthesis in cell-free systems. *Biotechnology and Bioengineering*. 2020;117:1137–1147. <https://doi.org/10.1002/bit.27249>