Published online 29 April 2016 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* ***W217–W225***
doi: 10.1093/nar/gkw342

MRE: a web tool to suggest foreign enzymes for the biosynthesis pathway design with competing endogenous reactions in mind

Hiroyuki Kuwahara†, Meshari Alazmi†, Xuefeng Cui and Xin Gao*

King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Thuwal, 23955, Saudi Arabia

Received February 05, 2016; Revised April 17, 2016; Accepted April 18, 2016

ABSTRACT

To rationally design a productive heterologous biosynthesis system, it is essential to consider the suitability of foreign reactions for the specific endogenous metabolic infrastructure of a host. We developed a novel web server, called MRE, which, for a given pair of starting and desired compounds in a given chassis organism, ranks biosynthesis routes from the perspective of the integration of new reactions into the endogenous metabolic system. For each promising heterologous biosynthesis pathway, MRE suggests actual enzymes for foreign metabolic reactions and generates information on competing endogenous reactions for the consumption of metabolites. These unique, chassis-centered features distinguish MRE from existing pathway design tools and allow synthetic biologists to evaluate the design of their biosynthesis systems from a different angle. By using biosynthesis of a range of high-value natural products as a case study, we show that MRE is an effective tool to guide the design and optimization of heterologous biosynthesis pathways. The URL of MRE is http://www.cbrc.kaust.edu.sa/mre/.

INTRODUCTION

Recent advances in genome editing and metabolic engineering enabled a precise construction of *de novo* biosynthesis pathways for high-value natural products (1,2). One important design decision to make for the engineering of heterologous biosynthesis systems is concerned with which foreign metabolic genes to introduce into a given host organism (3). Although this decision must be made based on multifaceted factors, a major one is the suitability of pathways for the endogenous metabolism of a host organism, in part because the efficacy of heterologous biosynthesis is affected by competing endogenous pathways (3–5). To address this point, we developed an open-access web server called MRE (Metabolic Route Explorer) that systematically searches for promising heterologous pathways by considering competing endogenous reactions in a given host organism (Figure 1).

There are various computational tools available to guide the design of heterologous pathways with a range of scopes and functions. Table 1 summarizes features of several heterologous pathway design tools. One key difference among these design tools is the specification of host organisms. Some tools, including BNICE (6), PredPath (7) and Metabolic tinker (8), were developed to explore pathways irrespective of the consideration for host organisms. Thus, these tools cannot assess the suitability of pathways for a specific context of the endogenous metabolic system of a host organism.

Some other tools, on the other hand, restrict the user to use one specific host organism. For example, pathway design tools based on flux balance analysis (FBA), such as XTMS (9), DESHARKY (10), OptStrain (11) and GEM-Path (12), are specific to the *Escherichia coli* chassis. While FBA-based tools tend to offer rich information to evaluate *de novo* pathways, they demand detailed knowledge of a given metabolic system with tight reaction-flux boundaries in order to identify meaningful steady-state flux distributions among a large number of candidate solutions. Clearly, such data are only available for well-studied organisms, and this may be a major reason why FBA-based tools focus on the pathway design in *E. coli*.

In contrast, FMM (13) and PHT (14) allow the user to select a host organism from a large set of choices. However, these tools do not use the chassis information to rank suitable biosynthesis pathways for a given endogenous metabolic system. Instead, with this information, PHT just reports which enzymes are not natively available in the host, whereas FMM suggests the introduction of foreign enzymes for some reactions in heterologous pathways.

*To whom correspondence should be addressed. Tel: +966 12 8080323; Fax: +966 12 8021241; Email: xin.gao@kaust.edu.sa
†These authors contributed equally to the paper as first authors.

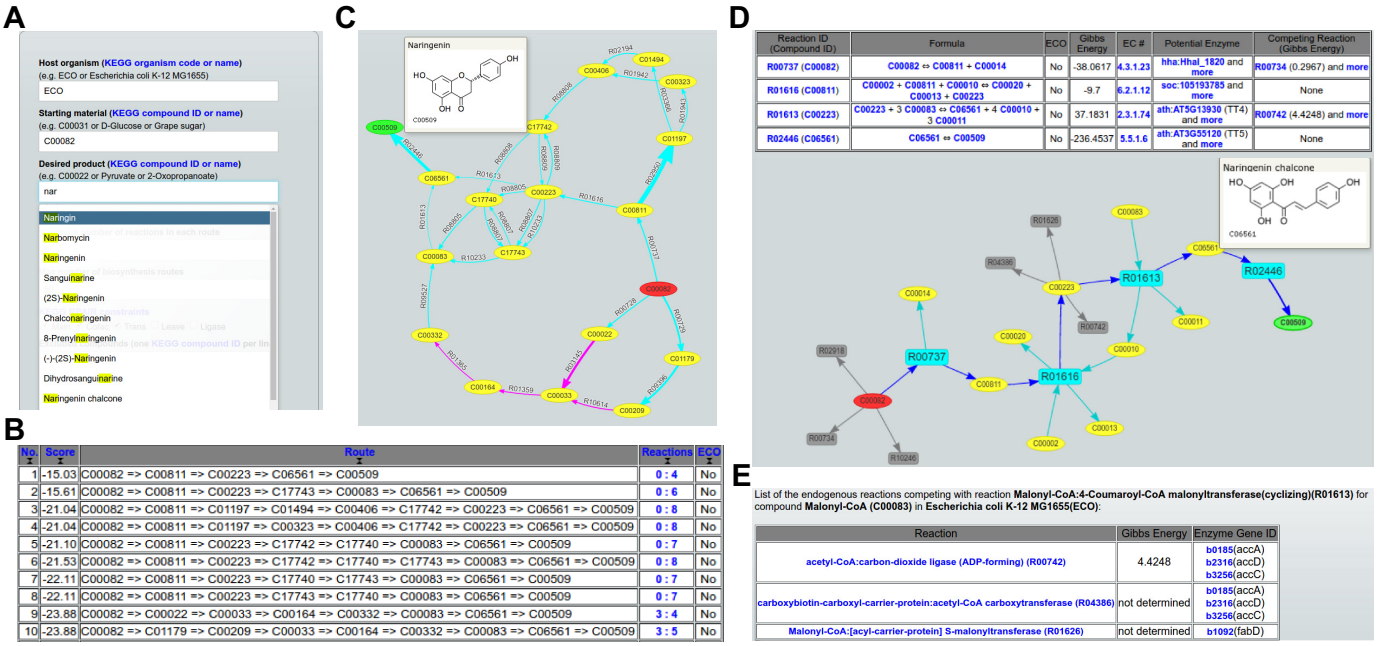


Figure 1. Typical user-interface pages of MRE. (**A**) The query input page. (**B**) The summary page for the top-ranked routes. (**C**) The page for a graph comprising the top 10 routes. (**D**) The page for pathway-level information. (**E**) The competing reaction information page.

Table 1. Feature summary of heterologous pathway design tools

Tool	Access	Chassis	Chemical transformation	Thermodynamic consideration	Ranking score	Information given for each pathway	(Ref.)
MRE	Open access web server	Many choices	Verified KEGG reactions	Boltzmann factor	Fraction of conversions via normalized Boltzmann weights	Required metabolites, EC numbers for enzymes, genes for foreign enzymes, reaction free energy, competing native reactions	–
FMM	Open access web server	Many choices	KEGG reactions	No	Number of reaction steps	EC numbers for enzymes, availability of each enzyme in various host organisms, suggestion for foreign enzymes	(13)
PHT	Open access web server	Many choices	KEGG reactions	No	Number of reaction steps	EC numbers for enzymes, local and global compound similarities for each reaction step	(14)
XTMS	Open access web server	*E. coli*	Predicted reactions	Favorability	Gene scores, reaction steps, toxicity, yield, Gibbs energy	Source compound for the retrosynthesis path, predicted reactions with EC numbers, genes for foreign enzymes, toxicity, production yield	(9)
Metabolic tinker	Open access web server	No host	RHEA reactions	Directionality, favorability	Net favorability	Possible reactions for each chemical transformation step and net favorability	(8)
PathPred	Open access web server	No host	Predicted reactions	No	Chemical similarity	Final compound of biodegradation, predicted intermediates and reactions, confidence for each predicted reaction	(7)
DESHARKY	Free download	*E. coli*	KEGG reactions	No	Growth rate	Source or target compound, EC numbers for enzymes, genes for some foreign enzymes, growth rate reduction measures	(10)
BNICE	Closed access	No host	Predicted reactions	No	No pathway ranking	3-level EC number for each predicted chemical transformation	(6)

Another main feature difference is the basis for chemical transformation of intermediate precursors that forms metabolic routes. While tools such as FMM, DESHARKY and Metabolic tinker specify chemical transformation using metabolic reaction sets from databases [e.g., KEGG (15) and RHEA (16)], other tools, including BNICE, PredPath and XTMS, predict some generalized chemical transformation rules using such curated reaction sets and apply them to expand potentially feasible metabolic routes.

Unlike these existing tools, MRE focuses on the suggestion of foreign enzymes with well-characterized activities for promising heterologous pathways by taking into account the effects of the existing, endogenous metabolic infrastructure of a host organism. To find promising biosynthesis routes from a large number of potential candidates, thermodynamic data offer useful information (17). Some existing pathway design tools, such as Metabolic tinker and XTMS, use thermodynamic data to constrain the reaction directionality or to rank pathways based on their net favorability, which does not consider competing endogenous reactions. In contrast, MRE uses thermodynamic data to rank pathways in a host-dependent manner from the perspective of the integration of new reactions into the endogenous metabolic system. In order to suggest actual foreign enzymes for the design of heterologous biosynthesis pathways, MRE only considers verified reactions as metabolic parts. For each foreign reaction in a suggested heterologous pathway, MRE generates information about endogenous reactions competing for metabolites. Since one effective approach to increase the productivity is to attenuate or

eliminate competing reactions (4,5,18), MRE may also offer useful insights into how to debottleneck and optimize heterologous pathways.

MATERIALS AND METHODS

Data resources

MRE makes use of several data resources. The main ones are from the KEGG databases (15). KEGG lists around 4000 organisms, which MRE uses for the selection of a host organism. The KEGG COMPOUND database is used to identify metabolites, while the KEGG REACTION database and the ExPASy ENZYME database (19) are used to find metabolic reactions with verified activities. The eQuilibrator dataset (20) is used to obtain the reaction Gibbs energy in the standard 1M concentration setting. The KEGG RPAIR database (21) is used to restrict search space based on the relation between reactants and products. The KEGG GENES database is used for DNA sequence data for enzymatic genes, and the KEGG taxonomy mapping dataset is used to calculate taxonomic distances.

Function of MRE

To explore biosynthesis routes with MRE, the user specifies a host organism and a pair of the starting and target compounds. To increase its usability and to help the user specify organisms and compounds, MRE comes with an auto-completion feature. With advanced options, the user can override the default setting for the metabolic route search. These options include the maximum number of reaction steps (denoted by n), the number of top-ranked pathways to generate (denoted by K), and a list of compounds that are not considered as primary metabolic precursors in the search, which we call the *exclusion list*. By default, n and K are set to 8 and 50, respectively, while the exclusion list is based on the one from Metabolic tinker (8) and has 101 compounds that have high degrees of connectivity in its metabolic network graph, for example, water, ATP and ADP. This exclusion list can also be customized to have other compounds (e.g., CO_2). In addition, MRE allows the user to constrain the chemical transformation of precursors based on RPAIR types (e.g., main, cofac and trans). These filtering schemes to constrain possible chemical transformations were reported to increase the relevance of the *de novo* biosynthesis route suggestion (22). By default, MRE considers chemical transformations based on main, cofac and trans RPAIR types.

Based on the input query for biosynthesis requirements, MRE generates the top-K metabolic routes, and the main result page summarizes these routes. For each metabolic route, MRE highlights whether it is endogenous or heterologous to the host organism. For each foreign reaction in a heterologous biosynthesis route, MRE predicts which metabolites may not be available in the host, and it lists exogenous genes for the corresponding enzymatic activity and suggests a list of foreign genes based on a taxonomic similarity measure whose cDNA sequences can be downloaded in the FASTA format. It also shows a list of native reactions competing for a metabolic precursor with each foreign enzymatic reaction. MRE provides a mean to visualize a specific pathway with competing endogenous reactions as well as a graph aggregating top-ranked routes.

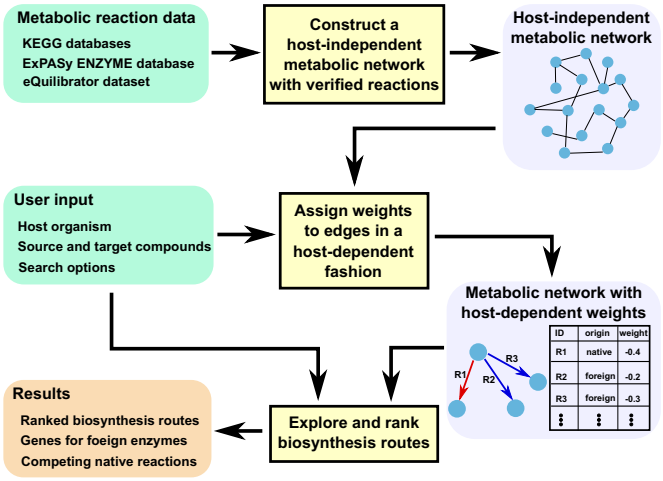


Figure 2. Workflow of MRE.

Workflow of MRE

Figure 2 depicts the workflow of MRE. Our tool first constructs a directed graph representing a host-independent metabolic network with verified reactions. This graph comprises all metabolic reactions with verified activities found in the data source, and it is built regardless of the choice of a host organism for a biosynthesis system. It next assigns weights to the edges in the graph in a host-dependent fashion by classifying which enzymatic reactions are native and foreign in the given host organism and by using the thermodynamic data. To search for biosynthesis routes from the starting material to the product in the host, MRE explores the host-independent metabolic network with the host-dependent weighting scheme exhaustively and generates top-K biosynthesis routes.

Host-independent metabolic network with verified reactions

To construct the host-independent metabolic network, we first identified metabolic reactions with verified activities. We categorized enzymatic reactions based on Enzyme Commission numbers (EC numbers) (23). Each EC reaction (i.e., a reaction class corresponding to each EC number) denotes a class of catalytic reactions with the same chemical transformation. To retrieve verified metabolic reactions with known enzymes, we filtered out reaction classes with partially qualified EC numbers as these partial EC reactions are unverified and can lead to misinterpretation of enzymatic activities (24). We also removed those EC reactions that do not contain any enzymes. With this filtering process, we identified 5389 complete EC reactions and 76 spontaneous reactions with verified activities.

We next estimated the standard reaction Gibbs energy $\Delta_r G'^{\circ}$ for each of these verified reactions using eQuilibrator with absolute temperature set to 298.15K. Each verified EC reaction is then split into two reactions: the forward reaction with the reaction Gibbs energy $\Delta_r G'^{\circ}$ and the

backward reaction with the reaction Gibbs energy $-\Delta_r G'^{\circ}$. Those EC reactions whose $\Delta_r G'^{\circ}$ could not be estimated were assigned the largest of the estimated values for both directions. We adopted this approach to take a conservative stance and avoid the suggestion of biosynthesis routes containing reactions with no thermodynamic information as much as possible. Using these reactions, we built a directed graph that models the transformation of metabolites where its vertices represent metabolites and its edges represent chemical transformations via verified metabolic reactions. Since this directed graph unifies all metabolic reactions with verified activities in the reaction databases, its structure is independent of the endogenous metabolic system of any host organism.

Host-dependent reaction weighting scheme

Using the user-specified host organism, we assigned weights to the edges of the directed graph. To assign the weight of each outgoing edge from a given compound node, we first assumed that this reaction was in the host organism and computed the probability of converting the precursor via this reaction over the competing native reactions. That is, by representing the competition for a metabolic precursor with endogenous reactions by a statistical mechanical model, we computed the probability of each reaction with $\Delta_r G'^{\circ}$ through the Boltzmann distribution. We next took the logarithm of this probability and assigned it as the weight of this outgoing edge. While this type of statistical mechanics modeling has been widely applied in the context of gene regulation to capture the promoter states (25–27), it is, to our knowledge, novel in the context of the biosynthesis system design. This weighting scheme depends on a host organism and models the competition for metabolic precursors with the endogenous reactions. Importantly, this competition-based weighting scheme can capture the effects of competing endogenous reactions on heterologous reactions, while a thermodynamic favorability-based weighting scheme cannot. This can make their weight assignments widely different from each other (as illustrated in Figure 3). A detailed description of this weighting scheme is given in Appendix.

Biosynthesis route search

Biosynthesis pathways of interest are often those that transform a higher fraction of a starting material to a target product. One heuristic to rank pathways based on this productivity criterion is the net favorability of pathways. At a first glance, the net thermodynamic favorability can be seen as a good measure to rank pathways based on this criterion. However, this measure can only quantify the ratio of the target concentration to the source concentration at equilibrium, which may not correspond well with the true picture of the titer of the target product, especially when a given pathway has strong competing reactions and the equilibrium concentration of the starting material is substantially lowered.

As described in the previous section, our reaction weighting scheme is based on the logarithm of normalized Boltzmann weights. Unlike thermodynamic favorability measure,

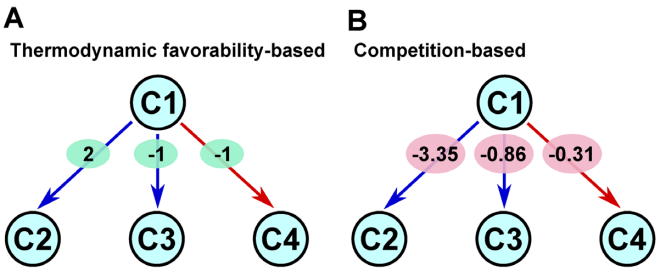


Figure 3. An illustration of differences between the thermodynamic favorability-based weighting scheme and our competition-based weighting scheme. Nodes are metabolites and edges are metabolic conversions via reactions. Red edges indicate native reactions, while blue edges indicate foreign reactions. (**A**) The thermodynamic favorability based approach. The value within a green oval for each edge represents the weight $\Delta_r G'^{\circ}/RT$ where R is the gas constant and T is the absolute temperature. (**B**) The competition-based approach. For each edge, the value within a pink oval represents its weight. With this scheme, edges with the same $\Delta_r G'^{\circ}$ value can have different weights in a host-dependent fashion. For example, the weight of C1 → C3 is ln $[e^1/(1 + e^1 + e^1)]$, while that of C1 → C4 is ln $[e^1/(1 + e^1)]$.

this estimates a fraction of a given precursor that is converted into next intermediate metabolites. Thus, a pathway score based on the sum of all reaction weights in a given pathway can characterize the lower bound of a fraction of starting material that is converted into the product through this pathway, and this score can capture the productivity of each pathway more appropriately.

Given the metabolic network graph with host-dependent weights, MRE exhaustively searches for biosynthesis paths from the given starting material to the given product and generates top-K metabolic routes, each of which has at most n reaction steps. In this search, the compounds in the exclusion list are not considered as intermediate precursors of the product. To rank routes, MRE computes their scores by summing all reaction weights in each route and keeps K routes with the highest scores (as illustrated in Figure 4). MRE transforms the metabolic route search problem into a classical computer science problem known as K-shortest loopless path problem (28) and uses an efficient algorithm to solve it. The core part of the search was implemented in C++.

RESULTS

Computational performance analysis in the runtime environment

To evaluate the computational performance of MRE, we measured its processing time in the runtime environment. To this end, we randomly selected 1000 reachable pairs of source and target compounds. With the setting of the largest reaction step size and the largest number of top-ranked pathways (i.e., $n = 20$ and $K = 500$), it took less than 10 s for MRE to exhaustively explore routes and process queries on average. In 95% of the samples, the processing time was less than 20 s, and even in the worst case, it was just less than 30 s. With the default setting (i.e., $n = 8$ and $K = 50$), the processing time was at most 1.36 s. Thus, we expect that the exhaustive pathway search employed in MRE will not compromise the user experience based on its processing time.

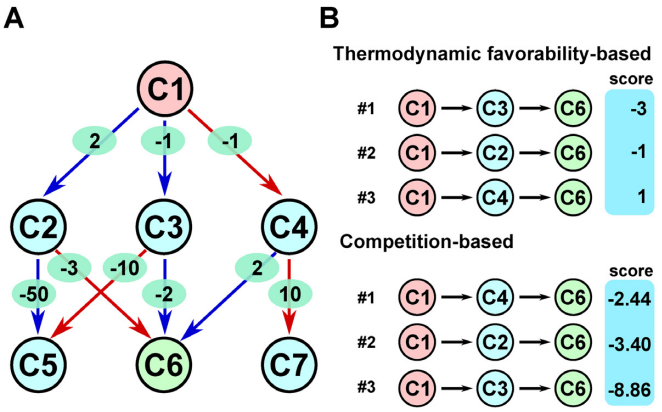


Figure 4. An illustrative example to show differences in ranking outcomes between the thermodynamic favorability based approach and our new resource utilization competition-based approach. (**A**) A simplified metabolic network. Nodes are metabolites and edges are metabolic conversions. Red edges indicate native reactions, while blue edges indicate foreign reactions. The value within a green oval for each edge indicates $\Delta_r G'^{\circ}/RT$ where R is the gas constant and T is the absolute temperature. Here, compound C1 is the starting metabolite, and compound C6 is the target product. There are three routes for this biosynthesis. (**B**) Ranking of the three biosynthesis routes with the thermodynamic favorability approach (the lower the score, the better) and our competition-based approach (the higher the score, the better). For example, the score of C1 → C4 → C6 is $-1 + 2 = 1$ with the former, whereas it is $\ln [e^1/(1 + e^1)] + \ln [e^{-2}/(1 + e^{-2} + e^{-10})] = -2.44$ with the latter.

Case study

As a case study, we applied MRE to search for pathways for various biosynthesis specifications using either *E. coli* K-12 MG1655 or *Saccharomyces cerevisiae* as the host organism. Table 2 summarizes the top-ranked heterologous pathways that MRE discovered. This shows that, in biosynthesis of a range of high-value natural products, MRE was able to identify pathways that are known to be productive. We also analyzed the results by comparing them with those from four open-access web servers that can design heterologous biosynthesis pathways, namely, FMM (13), Metabolic tinker (8), PHT (14) and XTMS (9). To explore biosynthesis pathways with these tools, we used their default configurations.

Biosynthesis of naringenin. Naringenin is a plant secondary metabolite, which is reported to have various health benefits (37), including high antioxidant capacities (38) and significant antiviral effects on the hepatitis C virus (39). Owing to inefficiencies in the production of naringenin from natural plant sources, metabolic engineering to have an efficient microbial synthesis of this high-value natural product is thought to be a commercially viable alternative (29,40).

In this analysis, we selected L-tyrosine (KEGG compound ID: C00082), an aromatic non-essential amino acid, as the starting material since a state-of-the-art heterologous naringenin production from L-tyrosine in an *E. coli* strain is known (see Figure 5A). This heterologous biosynthesis route comprises four foreign enzymatic reactions. To analyze the performance of MRE in comparison with other tools, we applied two open-access biosynthesis pathway web servers, Metabolic tinker (8) and XTMS (9). Since these two recently developed tools also rely on reaction thermodynamic data for their pathway ranking, we can also analyze the effects on our competition-based ranking scheme.

Given this biosynthesis requirement, Metabolic tinker and PHT were not able to find any pathways, while XTMS generated a predicted pathway with hypothetical reactions as its top-ranked candidate. In contrast, the top-ranked route from MRE and FMM was identical to the state of the art. The pathway information given by MRE indicates that the third reaction in the pathway, which transforms p-coumaroyl-CoA into naringenin chalcone, is a bottleneck and competes for the availability of cofactor malonyl-CoA with a more favorable native reaction involved in the fatty acid biosynthesis in the *E. coli* host (Figure 5B). This suggests that an increase in the concentration of malonyl-CoA or the inhibition of the fatty acid biosynthesis could enhance the productivity of this naringenin biosynthesis pathway. Indeed, previous studies demonstrated that both an increase in the availability of malonyl-CoA in the host and a decrease in the activities in the fatty acid pathway can increase the naringenin titers (29,41). While FMM was also able to identify the heterologous naringenin biosynthesis pathway that MRE found, the pathway information given by FMM was not helpful to find an optimization target as FMM does not have a feature to quantify the effects of competing reactions in the host.

Production of value-added chemicals from glycerol. Glycerol is a readily available and relatively inexpensive chemical compound that can be generated in large amounts as a byproduct of biodiesel and bioethanol production processes (42,43). Because of its economic viability and long-term sustainability, fermentative production of high-value materials from glycerol has gained much attention recently (44). By using glycerol as the starting material, we searched for pathways for the production of two value-added chemicals, 1,3-propanediol (1,3-PDO), a commodity chemical mainly used to make polyester fiber, and 1,2-propanediol (1,2-PDO), another high-demand commodity chemical used to make a wide range of products including antifreeze, thermoset plastics and cosmetics.

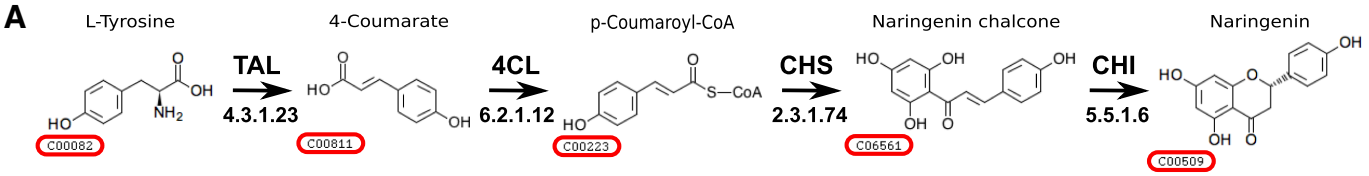
We first applied MRE to search for pathways for the production of 1,3-PDO in *E. coli* chassis. The top-ranked pathway that MRE identified is a known two-step heterologous pathway (42), which requires the introduction of a glycerol dehydratase gene and a 1,3-propanediol dehydrogenase gene in the host (Figure 6A). Since the first glycerol dehydratase reaction competes for the utilization of glycerol against several native reactions including glycerol kinase, MRE predicts that this can be a target for productivity optimization. Metabolic tinker and XTMS were not able to find any pathways for the 1,3-PDO production, whereas FMM and PHT found the same pathway that MRE identified.

We next applied MRE to search for pathways for the synthesis of R-1,2-PDO in the yeast chassis. We found that the top-ranked pathway (Figure 6B) was a known synthesis pathway for 1,2-PDO (31). In this pathway, glycerol is first converted to dihydroxyacetone phosphate (DHAP) via two native enzymatic reactions. Methylglyoxal synthase then transforms DHAP into methylglyoxal, which is,

Table 2. Top-ranked pathways identified by MRE for various biosynthesis specifications

Biosynthesis specification			Results of top-ranked pathway identified by MRE			Comparison with existing tools	
Source	Target	Host	Steps	Necessary foreign enzymes	Remark	Found a path[a]	Match with MRE[b]
L-tyrosine (C00082)	naringenin (C00509)	*E. coli* (ECO)	4	4.3.1.23, 6.2.1.12, 2.3.1.74, 5.5.1.6	Recovered a known route (29) as the top route	FMM, XTMS	FMM
glycerol (C00116)	1,3-PDO (C02457)	*E. coli* (ECO)	2	4.2.1.30, 1.1.1.202	Recovered a known route (30) as the top route	FMM, PHT	FMM, PHT
glycerol (C00116)	R-1,2-PDO (C02912)	yeast (SCE)	5	4.2.3.3, 1.1.1.79, 1.1.1.77	Recovered a known route (31) as the top route	MT	MT
acetyl-CoA (C00024)	artemisinic acid (C20309)	yeast (SCE)	10	2.5.1.92, 4.2.3.50, 4.2.3.82, 4.2.3.24, 1.14.13.158	Recovered a known route (32), and predicted better ones	none	none
L-tyrosine (C00082)	resveratrol (C03582)	*E. coli* (ECO)	3	4.3.1.23, 6.2.1.12, 2.3.1.95	Recovered a known route (33) as the top route	FMM	FMM
D-xylose (C00181)	xylitol (C00379)	*E. coli* (ECO)	2	1.1.1.21, 1.1.1.307	Recovered two known routes (34) as the top routes	FMM, PHT	FMM, PHT
PRPP (C00119)	histidine (C00135)	*E. coli* (ECO)	8	2.6.1.27	Predicted better and shorter routes than a known native route (35)	FMM, MT	none
chorismate (C00251)	tryptophan (C00078)	yeast (SCE)	5	none	Predicted the native route (36) as the best, and found shorter routes	FMM, MT, PHT	FMM

For each biosynthesis specification, the source and target compounds are specified in KEGG ID, and the host organism is in KEGG organism code. For each pathway, the number of reaction steps and the necessary foreign enzymes(in EC number) are specified. Comparison with FMM (13), Metabolic tinker (MT) (8), PHT (14) and XTMS (9) is also shown. For each tool, its default setting was used, except for the configuration of a pathway length, which was set to accommodate known pathways.
[a]Tools that have identified at least one path for a given biosynthesis specification.
[b]Tools whose top-ranked pathway is the same as the top-ranked one from MRE.



B **Pathway Number 1:** C00082 (R00737) C00811 (R01616) C00223 (R01613) C06561 (R02446) C00509
Download all potential cDNA gene sequences in fasta format for this pathway

Reaction ID (Compound ID)	Formula	ECO	Gibbs Energy	EC #	Potential Enzyme	Competing Reaction (Gibbs Energy)
R00737 (C00082)	C00082 ⇔ C00811 + C00014	No	-38.0617	4.3.1.23	hha:Hhal_1820 and more	R00734 (0.2967) and more
R01616 (C00811)	C00002 + C00811 + C00010 ⇔ C00020 + C00013 + C00223	No	-9.7	6.2.1.12	soc:105193785 and more	None
R01613 (C00223)	C00223 + 3 C00083 ⇔ C06561 + 4 C00010 + 3 C00011	No	37.1831	2.3.1.74	ath:AT5G13930 (TT4) and more	R00742 (4.4248) and more
R02446 (C06561)	C06561 ⇔ C00509	No	-236.4537	5.5.1.6	ath:AT3G55120 (TT5) and more	None

Figure 5. A heterologous biosynthesis pathway to produce naringenin from L-tyrosine in an *E. coli* host. (**A**) The structure of an experimentally derived biosynthesis pathway (29). The EC number for each reaction is indicated below the arrow. The KEGG compound ID of each metabolite is framed in red. The abbreviations are: tyrosine ammonia lyase (TAL); 4-coumarate:CoA ligase (4CL); chalcone synthase (CHS) and chalcone isomerase (CHI). (**B**) The information of the top-ranked biosynthesis pathway in MRE.

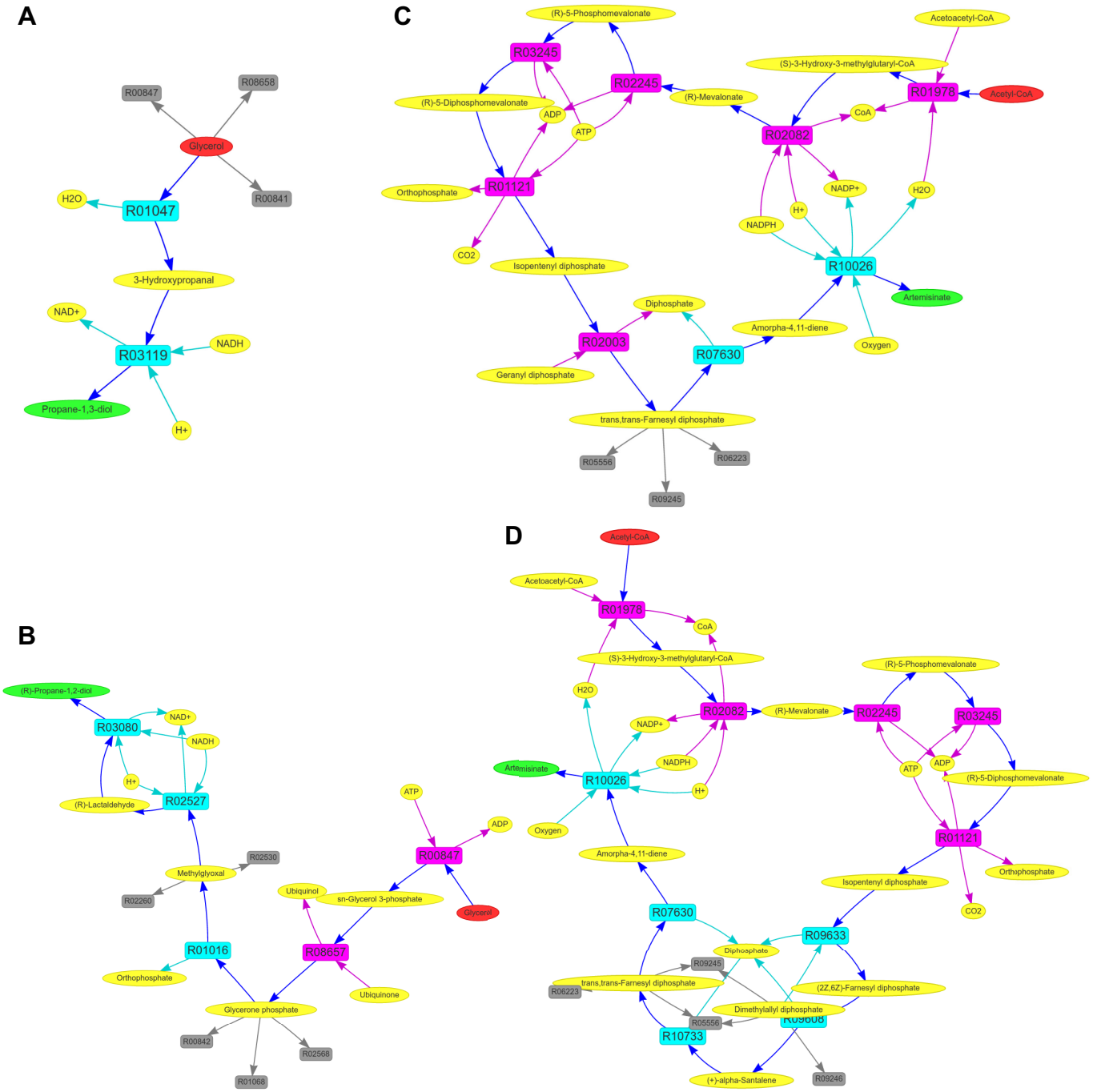


Figure 6. Pathway-level graphs generated in MRE. (**A**) The top-ranked pathway for the production of 1,3-PDO from glycerol in *E. coli*. (**B**) The top-ranked pathway for the production of R-1,2-PDO from glycerol in yeast. (**C**) A known pathway and (**D**) top-ranked pathway for the production of artemisinic acid from acetyl-CoA in yeast. In these graphs, oval nodes represent compounds, while box nodes represent reactions. For compound nodes, red nodes are the starting material, green nodes are the target products and yellow nodes are other compounds. For reaction nodes, purple nodes are native reactions, cyan nodes are foreign reactions and gray nodes are competing native reactions.

in turn, converted into (R)-lactaldehyde. Finally, lactaldehyde reductase is used to produce R-1,2-PDO from (R)-lactaldehyde. FMM and PHT were not able to find any pathways that convert glycerol into R-1,2-PDO, whereas Metabolic tinker identified the same pathway that MRE found as the top-ranked one. Since XTMS focuses on the *E. coli* chassis, we applied this tool to search for heterologous R-1,2-PDO production pathways in *E. coli*; however, no pathways were found.

Production of artemisinic acid. Artemisinic acid is an intermediate precursor for antimalaria drug artemisinin (32,45), and its production is often celebrated as one of the early success stories of the combination of metabolic engineer-

ing and synthetic biology (46). This engineered biosynthesis pathway utilizes the endogenous mevalonate pathway in budding yeast to transform acetyl-CoA into farnesyl pyrophosphate (FPP), which is then converted into artemisinic acid with heterologous amorphadiene synthase and three-step oxidation reactions (32,45).

To see if MRE could recover this engineered pathway, we applied it to explore pathways for the production of artemisinic acid from acetyl-CoA in yeast. We found that one of the top-ranked pathways that MRE generated was this known heterologous pathway (Figure 6C). Interestingly, the pathway that MRE identified as the top candidate (Figure 6D) was slightly different from the previously engineered pathway. The difference lies in how isopentenyl pyrophosphate (IPP) is converted into farnesyl pyrophosphate (FPP). In the top-ranked path, IPP is first converted into (2Z,6Z)-farnesyl diphosphate (Z,Z-FPP). This route is chosen because IPP is a precursor of a thermodynamically highly favorable native reaction, and the conversion reaction from IPP to Z,Z-FPP is much more favorable than that from IPP to FPP, enabling a higher fraction of IPP to be utilized in the route. By using Z,Z-FPP as the precursor, this route introduces three foreign carbon-oxygen lyases to form FPP. FMM, Metabolic tinker and PHT were not able to find any pathways. XTMS found a partial pathway that converts FPP into artemisinic acid, albeit it is for the *E. coli* chassis.

DISCUSSION

In this paper, we introduced MRE, an open-access biosynthesis design tool, that searches for promising metabolic routes for a given biosynthesis specification and suggests exogenous enzymes for heterologous biosynthesis pathways based on the infrastructure of an endogenous metabolic system. A main limitation of MRE is its reliance on the data sources (mainly KEGG) to mine verified metabolic reactions and to search for biosynthesis routes based on them. Indeed, while painstaking effort has resulted in a large collection of annotated metabolic reaction data, among the 9910 reactions in the KEGG REACTION database (Release 76.0), we found 1272 with no EC numbers, 1079 with partial EC numbers and 2170 with no annotations for associated genes. While this deficiency can prevent MRE from finding promising biosynthesis pathways, we expect the number of verified reactions in KEGG to increase over time and this issue to be alleviated eventually. At the same time, we are considering an option to also integrate other metabolic reaction databases such as Rhea (16) in a future release.

Several existing tools took an approach to expand a list of metabolic parts in hand by defining specific transformation rules (6,7,9), albeit such rules can be subjective (47). To design biosynthesis systems, this approach relies on the prediction of metabolic parts with specific metabolic activities, which may or may not exist. Thus, the design of biosynthesis systems via this top-down approach may require the *de novo* design of unnatural proteins to achieve specific metabolic activities. MRE was, on the other hand, developed to suggest actual enzymes for heterologous pathways. Thus, it takes a complementary, bottom-up approach in which a biosynthesis system is designed by using well-characterized metabolic parts. To this end, we made a conscious decision to use only verified reactions.

Here, by using the biosynthesis of a range of high-value natural products as a case study, we have shown that MRE can suggest promising heterologous biosynthesis pathways and provide useful information to pinpoint bottlenecks of pathways. In summary, with the host-dependent competition-based pathway ranking scheme along with the suggestion of foreign enzymes with competing endogenous reactions, MRE is expected to offer novel insights into the design and optimization of heterologous biosynthesis systems.

ACKNOWLEDGEMENTS

We thank Elad Noor for his help with eQuilibrator. This research made use of the resources of the computer clusters at KAUST.

FUNDING

King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) [URF/1/1976-04]. Funding for open access charge: King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) [URF/1/1976-04].
Conflict of interest statement. None declared.

REFERENCES

1. Luo,Y., Li,B.-Z., Liu,D., Zhang,L., Chen,Y., Jia,B., Zeng,B.-X., Zhao,H. and Yuan,Y.-J. (2015) Engineered biosynthesis of natural products in heterologous hosts. *Chem. Soc. Rev.*, **44**, 5265–5290.
2. Jakočiūnas,T., Jensen,M.K. and Keasling,J.D. (2016) CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.*, **34**, 44–59.
3. Prather,K.L.J. and Martin,C.H. (2008) *De novo* biosynthetic pathways: rational design of microbial chemical factories. *Curr. Opin. Biotechnol.*, **19**, 468–474.
4. Solomon,K.V., Moon,T.S., Ma,B., Sanders,T.M. and Prather,K. L.J. (2013) Tuning primary metabolism for heterologous pathway productivity. *ACS Synth. Biol.*, **2**, 126–135.
5. Ida,K., Ishii,J., Matsuda,F., Kondo,T. and Kondo,A. (2015) Eliminating the isoleucine biosynthetic pathway to reduce competitive carbon outflow during isobutanol production by *Saccharomyces cerevisiae*. *Microb. Cell Fact*, **14**, 62.
6. Hatzimanikatis,V., Li,C., Ionita,J.A., Henry,C.S., Jankowski,M.D. and Broadbelt,L.J. (2005) Exploring the diversity of complex metabolic networks. *Bioinformatics*, **21**, 1603–1609.
7. Moriya,Y., Shigemizu,D., Hattori,M., Tokimatsu,T., Kotera,M., Goto,S. and Kanehisa,M. (2010) PathPred: an enzyme-catalyzed metabolic pathway prediction server. *Nucleic Acids Res.*, **38**, W138–W143.
8. McClymont,K. and Soyer,O.S. (2013) Metabolic tinker: an online tool for guiding the design of synthetic metabolic pathways. *Nucleic Acids Res.*, **41**, e113.
9. Carbonell,P., Parutto,P., Herisson,J., Pandit,S.B. and Faulon,J.-L. (2014) XTMS: pathway design in an eXTended metabolic space. *Nucleic Acids Res.*, **42**, W389–W394.
10. Rodrigo,G., Carrera,J., Prather,K.J. and Jaramillo,A. (2008) DESHARKY: automatic design of metabolic pathways for optimal cell growth. *Bioinformatics*, **24**, 2554–2556.
11. Pharkya,P., Burgard,A.P. and Maranas,C.D. (2004) OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.*, **14**, 2367–2376.
12. Campodonico,M.A., Andrews,B.A., Asenjo,J.A., Palsson,B.O. and Feist,A.M. (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.

13. Chou,C.-H., Chang,W.-C., Chiu,C.-M., Huang,C.-C. and Huang,H.-D. (2009) FMM: a web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Res.*, **37**(suppl 2), W129–W134.
14. Rahman,S.A., Advani,P., Schunk,R., Schrader,R. and Schomburg,D. (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, **21**, 1189–1193.
15. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
16. Morgat,A., Axelsen,K.B., Lombardot,T., Alcántara,R., Aimo,L., Zerara,M., Niknejad,A., Belda,E., Hyka-Nouspikel,N., Coudert,E. *et al.* (2014) Updates in Rhea—a manually curated resource of biochemical reactions. *Nucleic Acids Res.*, **43**, D459–D464.
17. Ataman,M. and Hatzimanikatis,V. (2015) Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Curr. Opin. Biotechnol.*, **36**, 176–182.
18. Avalos,J.L., Fink,G.R. and Stephanopoulos,G. (2013) Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.*, **31**, 335–341.
19. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
20. Flamholz,A., Noor,E., Bar-Even,A. and Milo,R. (2012) eQuilibrator–the biochemical thermodynamics calculator. *Nucleic Acids Res.*, **40**, D770–D775.
21. Shimizu,Y., Hattori,M., Goto,S. and Kanehisa,M. (2008) Generalized reaction patterns for prediction of unknown enzymatic reactions. *Genome Inform.*, **20**, 299.
22. Faust,K., Croes,D. and van Helden,J. (2009) Metabolic pathfinding using RPAIR annotation. *J. Mol. Biol.*, **388**, 390–414.
23. Webb,E.C. (1992) *Enzyme nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes*, Academic Press, San Diego.
24. Green,M. and Karp,P. (2005) Genome annotation errors in pathway databases due to semantic ambiguity in partial EC numbers. *Nucleic Acids Res.*, **33**, 4035–4039.
25. Ackers,G.K., Johnson,A.D. and Shea,M.A. (1982) Quantitative model for gene regulation by λ phage repressor. *Proc. Natl. Acad. Sci. U.S.A.*, **79**, 1129–1133.
26. Arkin,A., Ross,J. and McAdams,H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, **149**, 1633–1648.
27. Kuwahara,H., Myers,C.J. and Samoilov,M.S. (2010) Temperature control of fimbriation circuit switch in uropathogenic *Escherichia coli*: quantitative analysis via automated model abstraction. *PLoS Comput. Biol.*, **6**, e1000723.
28. Yen,J.Y. (1971) Finding the K shortest loopless paths in a network. *Manag. Sci*, **17**, 712–716.
29. Santos,C.N.S., Koffas,M. and Stephanopoulos,G. (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab. Eng.*, **13**, 392–400.
30. Tang,X., Tan,Y., Zhu,H., Zhao,K. and Shen,W. (2009) Microbial conversion of glycerol to 1,3-propanediol by an engineered strain of *Escherichia coli*. *Appl. Environ. Microbiol.*, **75**, 1628–1634.
31. Jeon,E., Lee,S., Kim,D., Yoon,H., Oh,M., Park,C. and Lee,J. (2009) Development of a *Saccharomyces cerevisiae* strain for the production of 1,2-propanediol by gene manipulation. *Enzyme Microb. Technol.*, **45**, 42–47.
32. Ro,D.-K., Paradise,E.M., Ouellet,M., Fisher,K.J., Newman,K.L., Ndungu,J.M., Ho,K.A., Eachus,R.A., Ham,T.S., Kirby,J. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940–943.
33. Mei,Y.-Z., Liu,R.-X., Wang,D.-P., Wang,X. and Dai,C.-C. (2015) Biocatalysis and biotransformation of resveratrol in microorganisms. *Biotechnol. Lett.*, **37**, 9–18.
34. Cirino,P.C., Chin,J.W. and Ingram,L.O. (2006) Engineering *Escherichia coli* for xylitol production from glucose-xylose mixtures. *Biotechnol. Bioeng.*, **95**, 1167–1176.
35. Alifano,P., Fani,R., Liò,P., Lazcano,A., Bazzicalupo,M., Carlomagno,M.S. and Bruni,C.B. (1996) Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol. Rev.*, **60**, 44–69.
36. Toyn,J.H., Gunyuzlu,P.L., White,W.H., Thompson,L.A. and Hollis,G.F. (2000) A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast*, **16**, 553–560.
37. Hollman,P.C. and Katan,M.B. (1998) Bioavailability and health effects of dietary flavonols in man. *Arch. Toxicol. Suppl.*, **20**, 237–248.
38. Cavia-Saiz,M., Busto,M.D., Pilar-Izquierdo,M.C., Ortega,N., Perez-Mateos,M. and Muñiz,P. (2010) Antioxidant properties, radical scavenging activity and biomolecule protection capacity of flavonoid naringenin and its glycoside naringin: a comparative study. *J. Sci. Food Agr.*, **90**, 1238–1244.
39. Nahmias,Y., Goldwasser,J., Casali,M., van Poll,D., Wakita,T., Chung,R.T. and Yarmush,M.L. (2008) Apolipoprotein B–dependent hepatitis C virus secretion is inhibited by the grapefruit flavonoid naringenin. *Hepatology*, **47**, 1437–1445.
40. Fowler,Z.L. and Koffas,M.A. (2009) Biosynthesis and biotechnological production of flavanones: current state and perspectives. *Appl. Microbiol. Biotechnol.*, **83**, 799–808.
41. Leonard,E., Yan,Y., Fowler,Z.L., Li,Z., Lim,C.-G., Lim,K.-H. and Koffas,M.A. (2008) Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. *Mol. Pharm.*, **5**, 257–265.
42. Yazdani,S.S. and Gonzalez,R. (2007) Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Curr. Opin. Biotechnol.*, **18**, 213–219.
43. Quispe,C.A., Coronado,C.J. and Carvalho,J.A. Jr (2013) Glycerol: Production, consumption, prices, characterization and new trends in combustion. *Renew Sustain. Energy Rev.*, **27**, 475–493.
44. Clomburg,J.M. and Gonzalez,R. (2013) Anaerobic fermentation of glycerol: a platform for renewable fuels and chemicals. *Trends Biotechnol.*, **31**, 20–28.
45. Paddon,C.J., Westfall,P.J., Pitera,D., Benjamin,K., Fisher,K., McPhee,D., Leavell,M., Tai,A., Main,A., Eng,D. *et al.* (2013) High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, **496**, 528–532.
46. Paddon,C.J. and Keasling,J.D. (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.*, **12**, 355–367.
47. Martin,C.H., Nielsen,D.R., Solomon,K.V. and Prather,K. L.J. (2009) Synthetic metabolism: engineering biology at the protein and pathway scales. *Chem. Biol.*, **16**, 277–286.

APPENDIX

MATHEMATICAL DESCRIPTION OF THE HOST-DEPENDENT REACTION WEIGHTING SCHEME

To derive a mathematical description of the weighting scheme, we consider a scenario in which to generate weights for edges in the reactions transforming precursor C. Here, let $\mathbf{R_N}$ be a set of native reactions that can transform C in a given host organism. For each reaction r that can transform C, we set $e^{-\Delta_r G'^{\circ}/RT}$ as its Boltzmann factor. Then, we define $f(r)$, the normalized Boltzmann factor for r, as follows:

$$f(r) = \frac{e^{-\Delta_r G'^{\circ}/RT}}{1 + e^{-\Delta_r G'^{\circ}/RT} + \sum_{r' \in \mathbf{R_N} \setminus \{r\}} e^{-\Delta_{r'} G'^{\circ}/RT}}, \qquad (1)$$

where R is the gas constant and T is the absolute temperature. That is, those reactions that are not in the host organism do not affect the calculation of the Boltzmann distribution. If $r \in \mathbf{R_N}$, then $f(r)$ is simply based on the Boltzmann distribution of the native reaction system transforming compound C. On the other hand, if $r \notin \mathbf{R_N}$, then $f(r)$ is based on the Boltzmann distribution of the reaction system that contains all native reactions transforming C and foreign reaction r. With this scheme, in our graph, every edge that transforms C in reaction r has the weight $\log f(r)$.