Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors

Milenko Tokic[1], Noushin Hadadi[1], Meric Ataman[1], Dario S. Neves[2], Birgitta E. Ebert[2], Lars M. Blank[2], Ljubisa Miskovic[1], Vassily Hatzimanikatis[1]

[1] Laboratory of Computational Systems Biotechnology (LCSB), Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland.

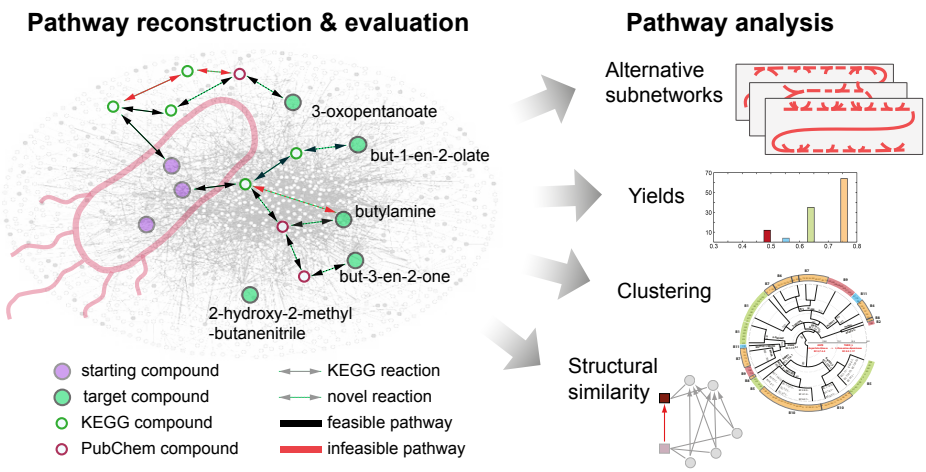
[2] Institute of Applied Microbiology (iAMB), Aachen Biology and Biotechnology (ABBt), RWTH Aachen University, D- 52056 Aachen, Germany.

Abstract

The limited supply of fossil fuels and the establishment of new environmental policies shifted research in industry and academia towards sustainable production of the 2nd generation of biofuels, with Methyl Ethyl Ketone (MEK) being one promising fuel candidate. MEK is a commercially valuable petrochemical with an extensive application as a solvent. However, as of today, a sustainable and economically viable production of MEK has not yet been achieved despite several attempts of introducing biosynthetic pathways in industrial microorganisms. We used BNICE.ch to discover all novel pathways around MEK. Out of 1'325 identified compounds connecting to MEK with one reaction step, we selected 3-oxopentanoate, but-3-en-2-one, but-1-en-2-olate, butylamine, and 2-hydroxy-2-methyl-butanenitrile for further study. We reconstructed 3'679'610 novel biosynthetic pathways towards these 5 compounds. We then embedded these pathways into the genome-scale model of *E. coli*, retaining a set of 18'925 most biologically viable ones based on their thermodynamic feasibilities and yields. For each novel reaction in the viable pathways, we proposed the most similar KEGG reactions, with their gene and protein sequences, as candidates for either a direct experimental implementation or as basis for enzyme engineering. Through pathway similarity analysis we classified the pathways and identified the enzymes and precursors that were indispensable for the production of the target molecules. The developments from this study enhance the potential of BNICE.ch for discovery, systematic evaluation, and analysis of novel pathways in future synthetic biology and metabolic engineering studies.

Keywords: Novel synthetic pathways, pathway feasibility, pathway similarity, Methyl Ethyl Ketone, 2-butanone, *E. coli*.

Graphical abstract



Introduction

Limited reserves of oil and natural gas and the environmental issues associated with their exploitation in the production of chemicals sparked off current developments of processes that can produce the same chemicals from renewable feedstocks. A fair amount of these efforts focuses on a sustainable production of the 2nd generation of biofuels.

Compared to the currently used fossil fuels and bioethanol, these 2nd generation biofuels should provide lower carbon emissions, higher energy density, and should be less corrosive to engines and distribution infrastructures. Currently, there is no fuel that satisfies all the above-mentioned criteria.[1] However, a large number of potential candidates has recently been proposed, such as n-butanol, isobutanol, 2-methyl-1-butanol or 3-methyl-1-butanol[2], C13 to C17 mixtures of alkanes and alkenes[3], fatty esters, and fatty alcohols[4].

For many of these chemicals, natural microbial producers are not known and novel biosynthetic pathways for their production are yet to be discovered.[5, 6] Even when production pathways for target chemicals are known, it is important to find alternatives in order to reduce cost and greenhouse emissions, and as well to avoid possible patent issues.

Computational tools are needed to assist in the design of novel biosynthetic pathways because they allow exhaustive generation of possible alternatives and evaluation of their properties and prospects for producing target chemicals[6]. For instance, computational tools can be used to assess the success of expression of a production pathway operating in one organism in another host organism. They can also be used to predict, prior to experimental pathway implementation, the yields across organisms of a particular pathway in producing a target molecule.

There are different computational tools for pathway prediction available in the literature.[7-17] An important class of these tools is based on the concept of generalized enzyme reaction rules, which were introduced by Hatzimanikatis and co-workers.[18, 19] These rules emulate the functions of enzymes, and they can be used to apply *in silico* biotransformations over a wide range of substrates.[6] Most of the implementations of this concept appear in the context of retrobiosynthesis, where the algorithm generates all possible pathways from a target compound toward desired precursors in an iterative backward manner.[5-7, 12, 14, 17-24]

In this study, we used the retrobiosynthesis framework of BNICE.ch[6, 7, 18-23] to explore the biotransformation space around Methyl Ethyl Ketone (MEK), also referred to as 2-butanone.[25] Besides acetone, MEK is the most commercially produced ketone with broad applications as a solvent for paints and adhesives and as a plastic welding agent.[26] MEK shows superior characteristics compared to the existing fuels in terms of its thermo-physical properties, increased combustion stability at low engine load, and cold boundary conditions, while decreasing particle emissions.[27] Since there is no known native producer of MEK, there were recent attempts to introduce biosynthetic pathways and produce this molecule in *E. coli*[28, 29] and yeast.[25] However, none of these attempts achieved commercially viable amounts of produced MEK. Alternatively, hybrid biochemical/chemical approaches that combine both fermentative and catalytic processes were proposed.[30, 31]

We used the BNICE.ch algorithm to generate a network of potential biochemical reactions around MEK, and we identified 159 KEGG[32, 33] and 1'166 PubChem[34, 35] compounds one reaction step away from MEK (Table S1 in the Supporting

Information). Out of the 1'325 compounds, 2-hydroxy-2-methyl-butanenitrile was the only compound with a known biotransformation to MEK. We chose this compound along with 3-oxopentanoate, but-3-en-2-one, but-1-en-2-olate and butylamine for further study. The latter four compounds were chosen based on their: (i) easy chemical conversion to MEK, e.g., 3-oxopentanoate spontaneously decarboxylates to MEK; and (ii) potential use as precursor metabolites to produce a range of other valuable chemicals.[36-38]

We have reconstructed all possible novel biosynthetic pathways (3'679'610 in total) up to a length of 4 reaction steps from the central carbon metabolites of *E. coli* toward the 5 compounds mentioned above. We evaluated the feasibility of these 3'679'610 pathways, and we further analyzed all 18'925 thermodynamically feasible pathways regarding their yields. We also identified metabolic subnetworks that were carrying fluxes when the optimal yields were attained. From this analysis, we determined the minimal sets of precursors and the common routes and enzymes for production of the target compounds.

Results and Discussion

Generated metabolic network around Methyl Ethyl Ketone

We used the retrobiosynthesis algorithm of BNICE.ch to iteratively reconstruct the biochemical network that contained all compounds and reactions that were up to five generations away from MEK (Figure 1). In the first iteration of the reconstruction procedure, we provided to BNICE.ch the initial set of compounds that contained 26 cofactors along with MEK (Table S2 in the Supporting Information). After five iterations, a total of 13'498 compounds were generated, and out of these, we found 749 in the KEGG[32, 33] and 13'414 in the PubChem[34, 35]

databases (Figure 1, panel A). These compounds were involved in 65'644 reactions, out of which 560 existed in the KEGG database and the remaining 65'084 were novel reactions (Figure 1, panel B). Out of 361*2 bidirectional generalized enzyme reaction rules of BNICE.ch, 369 (190 forward and 179 reverse) were required to generate the metabolic network around MEK with the size of 5 reaction steps.

Figure 1. Growth of the BNICE.ch generated metabolic network over 5 iterations. Compounds: KEGG (Panel A, blue) and PubChem (Panel A, red). Reactions: KEGG (Panel B, blue) and novel (Panel B, red).

In the first BNICE.ch iteration, there were 25 KEGG and 6 PubChem compounds connected through 48 novel reactions to MEK. After five iterations, BNICE.ch generated a total of 1'841 reactions that involved MEK. Among them, only one reaction was catalogued in the KEGG database (KEGG R09358). The generated reactions involved 1'325 carbon compounds that could be potentially used as MEK precursors. From these, we chose to study 2-hydroxy-2-methylbutanenitrile because it was the only KEGG compound connected to MEK through a KEGG reaction (KEGG R09358). We also selected four other compounds obtained in the first iteration: 3-oxopentanoate (KEGG compound), but-3-en-2-one (KEGG compound), butylamine (KEGG compound), and but-1-en-2-olate (PubChem compound).

Pathway reconstruction toward five target compounds

In the pathway reconstruction process, we used as starting compounds 157 metabolites selected from the generated network, which were identified as native *E. coli* metabolites using the latest *E. coli* genome-scale model iJO1366[39] (Table S3 in the Supporting Information). We performed an exhaustive pathway search on the generated metabolic network, and we reconstructed 3'679'610 pathways toward these five target compounds with pathway lengths ranging from 1 up to 4 reaction steps (Table 1). The reconstructed pathways consist of 37'448 reactions, i.e., 57% of the reactions reproduced from the BNICE.ch generated metabolic network.

More than 58% of the discovered pathways were toward butylamine. For but-1-en-2-olate, which was the only PubChem target compound, we discovered only 140'779 pathways, i.e., 3.8% of the reconstructed pathways (Table 1). From the 33 reconstructed pathways with the length of one reaction step, 28 were towards butylamine and none towards but-1-en-2-olate. The majority of reconstructed pathways (> 97%) had the length of four reaction steps. This result suggests that the biochemistry of enzymatic reactions favors smaller changes of a molecule structure over several steps, rather than a more radical change in one or a just a few steps.

Table 1. Reconstructed pathways toward five target compounds.

Target compounds	Reconstructed pathways	Reaction steps				Feasible pathways	
		1	2	3	4	FBA	TFA
3-Oxopentanoate	641'493	1	198	12'222	629'072	361'187	11'145
But-3-en-2-one	438'889	1	136	7'554	431'198	57'173	4'117
Butylamine	2'146'890	28	1'236	53'573	2'092'053	27'211	1'177
But-1-en-2-olate	140'779	0	53	2'905	137'821	30'689	1'826

2-Hydroxy-2-methylbutanenitrile	311’559	3	94	6’546	304’916	11’151	660
	3’679’610	33	1’717	82’800	3’595’060	487’411	18’925

Evaluation of reconstructed pathways

We performed a series of feasibility studies on the ~3.6 million generated pathways to assess their biological viability and performance, and we rejected the ones that did not pass the requirements (Methods). The feasibility of the pathways depends on the metabolic network of the chassis organism. Therefore, we embedded each of the reconstructed pathways in the *E. coli* genome-scale model iJO1366 to perform several feasibility tests. Unless stated otherwise, we applied the constraints from the *C1* set (Methods) for the subsequent analyses.

Flux balance analysis. We used FBA as a prescreening method to reject all the pathways that did not satisfy the mass balance constraints. More precisely, the FBA infeasible pathways were incompatible with the host organism as they required co-substrates that were absent in iJO1366. Out of all reconstructed pathways, 13.24% (487’411) were FBA feasible (Table 1). Though the largest number of reconstructed pathways were towards butylamine, only 1.27% (27’211) of these passed the FBA test. The number of FBA feasible pathways for 2-hydroxy-2-methylbutanenitrile was also low (3.59%). In contrast, more than 56% of pathways towards 3-oxopentanoate were FBA feasible.

Thermodynamics-based flux analysis. We used Thermodynamics-based Flux Analysis (TFA) to identify thermodynamically feasible pathways. Our analysis showed that TFA is a necessary step in the pathway evaluation process since a majority of FBA feasible pathways were in fact infeasible when subjected to the thermodynamic constrains. Only 18’925 pathways passed this test, i.e., 0.51% of

all generated pathways, or 3.88% of the FBA feasible pathways. The set of TFA feasible pathways involved 3'269 unique reactions.

Analogous to the FBA feasibility, the lowest TFA feasibility rate was for butylamine, i.e., 0.05% of reconstructed pathways were TFA feasible (Table 1). The highest rate of TFA feasible pathways was again for 3-oxopentanoate (1.74 %). The shortest TFA feasible pathways consisted of 2 reaction steps (21 pathways), whereas a majority of feasible pathways had length 4 (Table 2). All pathways contained novel reaction steps, and 19 pathways had only one novel reaction step (Table 2). All of these 19 pathways were towards but-3-en-2-one, and had as intermediates 2-acetolactate and acetoin. The final reaction step converting acetoin to but-3-en-2-one was novel for all of them.

Table 2. Number of known reaction steps versus all reaction steps. Pathways with one novel reaction step are marked in red. All shown pathways are TFA feasible.

		Reaction steps			Feasibility	
		2	3	4	TFA	
Number of known steps in a pathway	0	14	371	7'059	7'444	3-Oxopentanoate
	1		118	2'956	3'074	
	2			627	627	
	0	4	72	3'196	3'272	But-3-en-2-one
	1	1	13	703	717	
	2		2	110	112	
	3			16	16	
	0	2	35	974	1'011	Butylamine
	1		7	139	146	
	2			20	20	

	0	23	1’576	1’599	But-1-en-2-olate
	1	10	196	206	
	2		21	21	
	0	50	380	430	2-Hydroxy-2-methylbutanenitrile
	1	2	202	204	
	2		26	26	
		21 703	18’201	18’925	

Yield analysis. We used TFA to assess the production yield of the feasible pathways from glucose to the target compounds (S4 in the Supporting Information). We identified pathways for all target compounds that could operate without a loss of carbon from glucose. More than a half of the pathways toward 3-oxopentanoate (57%) could operate with the maximum theoretical yield of 0.774 g/g, i.e., 1Cmol/1Cmol (Figure 2). In contrast, only 27 out of 660 pathways toward 2-hydroxy-2-methylbutanenitrile (4%) could operate with the maximal theoretical yield of 0.66 g/g (S4 in the Supporting Information). The population of yields was grouped into several distinct sets rather than being more spread and continuous, e.g., eleven sets for 3-oxopentanoate, (S4 in the Supporting Information). Interestingly, a similar discrete pattern in pathway yields was observed for the production of mono-ethylene glycol in *Moorella thermoacetica* and *Clostridium ljungdahlii.*[40]

Analysis of alternative assumptions on reaction directionalities. Since the directionality of reactions in the network impacts yields, we investigated how

the set of alternative constraints *C2* (Methods) affected the yield distribution. The constraints from the C2 set imposed the directionalities from iJO1366 to some reactions that could operate in both directions with the constraints from the C1 set. As expected, these additional constraints reduced flexibility of the metabolic network. We observed that the distribution of yields was more spread compared to the one obtained using the *C1* constraints; the yields were in general lower, and some pathways even became infeasible (S5 in the Supporting Information). For example, there were three alternative pathways for the production of 3-oxopentanoate from acetate via two intermediate compounds: 2-ethylmalate and (3S)-3-hydroxypentanoate. The three alternative pathways had three different cofactor pairs in the final reaction step that converts (3S)-3-hydroxypentanoate to 3-oxopentanoate (Figure S1 in the Supporting Information). With the set of constraints *C1* applied, the three pathways had an identical maximal yield of 0.642 g/g. In contrast, with the set *C2* applied, the pathway with NADH/NAD cofactor pair in the final step had a yield of 0.537 g/g, the one with NADPH/NADP had a yield of 0.542 g/g, and the one with H_2O_2/H_20 had a yield of 0.495 g/g. These difference in yields are a consequence of the different costs of cofactor production upon adding supplementary constraints.

BridgIT analysis. For each novel reaction from the feasible pathways, we identified the most similar KEGG reaction whose gene and protein sequences were assigned to the novel reaction (Methods). The BridgIT results are available at http://lcsb-databases.epfl.ch/pathways/.

Analysis of non-core subnetworks capable of synthesizing target molecules

For each of the feasible pathways, we identified the active metabolic subnetworks containing the reactions required to carry fluxes to synthesize the corresponding target molecule (Methods). We then split the active metabolic subnetworks into the core metabolic network, which included central carbon metabolism pathways[41, 42], and the active non-core metabolic subnetwork (Figure 2, panel A, and Methods). In this way, we identified 57'139 active non-core subnetworks from the 18'925 TFA feasible pathways. On average, there were more than 3 alternative subnetworks per pathway due to the redundant topology of metabolism (Table 3).

Figure 2. Metabolic network representing the production of 3-oxopentanoate from glucose (panel A). Black lines: reactions pertaining to the core metabolic network. Red lines: reactions pertaining to the active non-core metabolic subnetwork. Green nodes: metabolites in the core metabolic network. Orange nodes: metabolites in the active non-core metabolic subnetwork. Yellow nodes: core precursors, i.e., metabolites that connect the core and active non-core subnetworks. Alternative pathways connecting ribose-5-phosphate, R5p, with 2-deoxy-D-ribose-1-phosphate, 2dr1p (panel B). Alternative pathways connecting the core metabolites with propanal, Ppal (panel C).

Table 3. Alternative non-core subnetworks for 5 target compounds together with their lumped reactions and precursors.

Target compounds	Feasible pathways	Alternative subnetworks	Unique lumps	Overlapping sets of precursors	Unique precursors
3-Oxopentanoate	11'145	35'013	4'517	281	40
But-3-en-2-one	4'117	10'162	1'762	126	32

Butylamine	1’177	4’259	1’791	102	30
But-1-en-2-olate	1’826	5’339	1’536	97	30
2-Hydroxy-2-methylbutanenitrile	660	2’420	794	37	17
	18’925	57’139	10’400		

Next, we computed a lumped reaction for each of the alternative subnetworks (Methods). Out of the 57’139 computed lumped reactions, only 10’400 were unique (Table 3). For the compound 3-oxopentanoate, we observed the largest diversity in alternative subnetworks per lumped reaction, where for 35’013 alternative subnetworks there were 4’517 unique lumped reactions, i.e., on average, more than 7 alternative subnetworks had the same lumped reactions. In contrast, we observed the smallest diversity for butylamin with more than two alternative subnetworks per lumped reaction (794 unique lumped reactions from 2’420 alternative subnetworks) (Table 3). For the five target compounds, there were, on average, more than 5 alternative subnetworks per lumped reaction. This result suggests that the overall chemistry and the cost to produce the corresponding target molecule are the same for many different pathways. An illustrative example of multiple pathways with the same lumped reaction is provided in Figure 2 in the Supporting Information.

Origins of diversity of alternative subnetworks. To better understand the diversity in alternative subnetworks, we performed an in-depth analysis of the two-step pathway from acetyl-CoA and propanal to 3-oxopentanoate, which was selected because it presented the largest number of alternative networks among all reconstructed pathways. The smallest alternative subnetwork of the 185 analyzed consisted of 14 enzymes, whereas the largest one comprised 22 enzymes (Table S6 in the Supporting Information). All subnetworks shared five

common enzymes: two enzymes from the production pathway converting propanal via (3S)-3-hydroxypentanoate to 3-oxopentanoate (with the BNICE.ch assigned third level Enzymatic Commission, EC, numbers[43] 2.3.3.- and 1.1.1.-), two enzymes involved in acetyl-CoA production (phosphopentomutase (deoxyribose), PPM2, and deoxyribose-phosphate aldolase, DRPA), and aldehyde dehydrogenase, ALDD3y, converting propionate (ppa) to propanal (Figure 2).

The multiplicity of ways to produce acetyl-CoA and propionate contributed to a large number of alternative subnetworks: there were 102 alternative ways of producing acetyl-CoA from ribose-5-phosphate (r5p) via 2-deoxy-D-ribose-1-phosphate (2dr1p) and 9 different ways of producing propionate (Figure 2, panels B and C).

There were two major routes to produce 2dr1p. In the first route, R5p is converted either to ribose-1-phosphate (in 31 alternatives) or to D-ribose (in 19 alternatives), which are intermediates in producing nucleosides such as adenosine, guanosine, inosine and uridine. These nucleosides are further converted to deoxyadenosine (Dad_2), deoxyguanosine (Dgsn) and deoxyuridine (Duri) that are ultimately phosphorylated to 2dr1p. In 26 of the 52 alternatives for the second route, R5p is converted to phosphoribosyl pyrophosphate (prpp), which is followed by a transfer of its phospho-ribose group to nucleotides such as AMP, GMP, IMP and UMP. These nucleotides are then converted to 2dr1p by downstream reaction steps. In the remaining alternatives for the second route, R5p is first converted to AMP in one reaction step, and then to 2dr1p via Dad_2 and Dgsn.

There were 9 alternative routes to produce propionate. In 4 of these, this compound was produced from pyruvate and succinate (Figure 2, panels A and

C), in 3 routes it was produced from aspartate (Figure 2, panel C), and in 2 routes it was produced from 3-phosphoglycerate and glutamate.

Core precursors of five target compounds. An abundant provision of precursor metabolites is crucial for an efficient production of target molecules.[44] Here, we defined as *core precursors* the metabolites that connect the core to the active non-core metabolic subnetworks (Figure 2, panel A). We analyzed the different combinations of core precursors that appeared in the alternative subnetworks. Our analysis revealed that the majority of subnetworks were connected to the core network through a limited number of core precursors. For example, 35'013 alternative subnetworks for the production of 3-oxopentanoate were connected to the core network by 281 different sets of core precursors (Table 3). In these 281 sets, there were only 40 unique core precursors. We ranked these sets based on their number of appearances in the alternative networks. The top ten sets appeared in 24'210 subnetworks, which represented 69% of all identified subnetworks for this compound (Table 4). Moreover, the metabolites from the top set (acetyl-CoA, propionyl-CoA, pyruvate, ribose-5-phosphate, and succinate) were the precursors in 8'510 (24.3%) subnetworks for 3-oxopentanoate (Table 4). Ribose-5-phosphate appeared in 9 out of the top ten sets, and it was a precursor in 32'237 (92%) 3-oxopentanoate producing subnetworks.

Table 4. Top ten core precursor combinations for the production of 3-oxopentanoate. Core precursors: acetate (Ac), acetyl-CoA (AcCoA), aspartate

(Asp-L), dihydroxyacetone phosphate (Dhap), propionyl-CoA (PpCoA), pyruvate (Pyr), ribose-5-phosphate (R5p), succinate (Succ), succinyl-CoA (SucCoA).

Ac	AcCoA	Asp-L	Dhap	PpCoA	Pyr	R5p	Succ	SucCoA	Number of subnetworks
	✔			✔	✔	✔	✔		8’510
		✔		✔		✔			5’409
				✔	✔	✔	✔		3’463
✔						✔		✔	1’344
					✔	✔	✔		1’049
	✔					✔	✔		965
	✔					✔		✔	956
				✔	✔		✔		915
	✔		✔			✔			834
			✔			✔			765
									24’210

Clustering of feasible pathways

The repeating occurrences of core precursors and lumped reactions in the alternative non-core subnetworks motivated us to identify common patterns in enzymes, core precursors and intermediate metabolites required to produce the target molecules. To this end, we performed two types of clustering on the test study of 115 feasible pathways from acetate to 3-oxopentanoate.

Clustering based on core precursors and byproducts of lumped reactions. We computed 242 lumped reactions corresponding to 115 pathways from the test study. For simplicity, we chose the first lumped reaction returned by the solver for each of the 115 pathways, and we discuss here the clustering based on core precursors and byproducts of the chosen 115 lumped reactions (Methods).

The main branching condition among the 115 pathways was the presence or absence of thioesters, such as AcCoA, in the set of core precursors (Figure 3). There were 56 pathways with CoA-related precursors and 59 pathways that did not require CoA. We further clustered the pathways from the former group subject to the presence of the precursors AcCoA (1 pathway), PpCoA (30 pathways), both AcCoA and PpCoA (6 pathways), and SucCoA (19 pathways), or the occurrence of the byproducts malonate (Maln) or CO_2. The pathways that did not require CoA were further clustered depending on if they had as precursors Formate (For) or DHAP (27 pathways) or not (32 pathways).

Figure 3. Clustering dendrogram of the 115 reconstructed pathways from acetate to 3-oxopentanoate and their respective yields (inset). Pathways were classified based on core precursors (red) and byproducts (green) of their lumped reactions. (R)-CoA denotes the group of thioesters.

Remarkably, the clustering based on core precursors and byproducts of lumped reactions also separated the pathways based on their yields (Figure 3, inset). For instance, pathways that had AcCoA, PpCoA, Dhap, and For as precursors had a maximal theoretical yield of 0.774 g/g. In contrast, pathways with 2-oxoglutarate (AKG) or SucCoA as precursors, and Maln as the byproduct, had the lowest yield (0.483 g/g) from the set of examined pathways.

The clustering also provided insight into the different chemistries behind the analyzed pathways. For most of the pathways, i.e., the ones classified in groups B1-2 and B4-10, there was a clear link between the core precursors and co-substrates of acetate in the first reaction step of the pathways (Figure 3). For

example, the pathways from the group B1 have a common first reaction step (EC 2.8.3.-) that converts acetate and 3-oxoadipyl-CoA to 3-oxoadipate (Figure 3). The clustering grouped these pathways together because SucCoA was the core precursor of 3-oxoadipyl-CoA through 3-oxoadipyl-CoA thiolase (3-OXCOAT). Moreover, 3-oxoadipate, a 6-carbon compound, was converted in downstream reaction steps to 3-oxopentanoate, a 5-carbon compound, and one molecule of CO_2 through 18 alternative routes. Similarly, in the single pathway of group B2 the co-substrate in the first reaction step was (S)-methylmalonyl-CoA, which was produced from SucCoa through methylmalonyl-CoA mutase (MMM). This enzyme, also known as sleeping beauty mutase, is a part of the pathway converting succinate to propionate in *E. coli*.[45] Malonate (Maln), a 2-carbon compound, was released in the first reaction step, which resulted in a low yield of this pathway (Figure 3).

Despite sharing the first reaction step in which acetate reacted with 2-oxoglutarate to create 2-hydroxybutane 1-2-4-tricarboxylate, the pathways from group B9 were split in two groups with different yields (Figure 3). These two groups differed in the sequences of reactions involved in the reduction of 2-hydroxybutane 1-2-4-tricarboxylate, a 7-carbon compound, to 3-oxopentanoate. In 11 pathways, the yield was 0.483 g/g due to a release of two CO_2 molecules, whereas in one pathway the yield was 0.644 g/g due to malate being created as a side-product and recycled back to the system.

Pathways from group B3 utilized different co-substrates, such as ATP and crotonoyl-CoA, along with acetate to produce acetaldehyde in the first reaction step. All these pathways shared a common novel reaction step with acetaldehyde and propionyl-CoA as substrates (EC 2.3.1.-).

Finally, group B11 contained the pathways with the intermediate 2-methylcitrate, which was produced from pyruvate (Pyr).

Clustering based on involved enzymes. The clustering based on the core precursors and byproducts provided an insight of the chemistry underlying the production of 3-oxopentanoate. However, lumped reactions hide the identity of the enzymes involved in the active non-core subnetworks. To find common enzyme routes for the production of 3-oxopentanoate, we performed a clustering based on the sets of enzymes forming the non-core subnetworks (Methods).

Five enzymes, AMP nucleosidase (AMPN), 5'-nucleotidase (NTD6), purine-nucleoside phosphorylase (PUNP2), PPM2 and DRPA, which participated in the production of acetaldehyde from R5p, were present in all routes from 3-oxopentanoate to acetate (Figure 2, panels A and C). The clustering separated pathways depending on if they contained or not a sequence of 6 enzymes starting with aspartate kinase (ASPK) and ending with L-threonine deaminase (THRD_L), whose product 2-oxobutanoate was converted downstream to 3-oxopentanoate (Figure 4). The groups contained 47 and 68 pathways, respectively.

Figure 4. Clustering dendrogram of 115 reconstructed pathways from acetate to 3-oxopentanoate. Pathways were clustered based on enzymes involved in the active non-core metabolic subnetworks.

Both groups were further clustered based on a set of enzymes required to produce deoxyadenosine and the downstream metabolite acetaldehyde (Figure

4). The first subgroup of enzymes, i.e. ribonucleoside-diphosphate reductase (RNDR1), deoxyadenylate kinase (DADK) and NTD6, converted adp to deoxyadenosine. In the second subgroup, atp was transferred to deoxyadenosine via ribonucleoside-triphosphate reductase (RNTR1c2), nucleoside triphosphate pyrophosphorylase (NTPP5) and NTD6. Then, for both subgroups, deoxyadenosine was converted to 2-deoxy-D-ribose 5-phosphate (2dr5p) that was further transformed to acetaldehyde via DRPA (Figure 2).

The clustering based on enzymes allowed us to identify enzymatic routes corresponding to different yields (Figure 4, and Figure 3 inset). For example, all pathways that include ASPK and novel reaction steps with BNICE.ch third level EC assignments, such as 1.13.11-, 1.2.1-, would provide the maximal theoretical yield of 0.774 g/g (Figure 4). Similarly, pathways that contained ALDD3Y, methylisocitrate lyase (MCITL2), and RNTR1C2, but not 3-OXCOAT and ASPK, would also provide the maximal theoretical yield. In contrast, the clustering also permitted us to identify key enzymes participating in pathways with a reduced yield. For example, pathways that contained 3-OXCOAT had a yield of 0.644 g/g.

Furthermore, the clustering based on enzymes allowed us to clarify the link between the precursors and the corresponding sequence of enzymes that needed to be active for producing the target molecule. For example, pathways from group B1, which had SucCoA as a core precursor and CO_2 as a byproduct, had the common reaction step 3-OXCOAT (Figure 4). Similarly, all pathways from group B4 with core precursors PpCoA and AcCoA contained ALDD3Y.

Ranking of biosynthetic pathways and recommendations

We ranked the corresponding feasible pathways according to their yield, number

of reaction steps and enzymes that could be directly implemented or needed to be engineered (Methods). This way, we obtained the top candidate pathways for each of the target molecules that were likely to produce these compounds with economically viable yields. The top candidates were visualized and can be consulted at http://lcsb-databases.epfl.ch/pathways/.

Methods

We employed the BNICE.ch framework[6, 7, 18-23] to generate biosynthetic pathways towards 5 precursors of Methyl Ethyl Ketone: 3-oxopentanoate, 2-hydroxy-2-methylbutanenitrile, but-3-en-2-one, but-1-en-2-olate and butylamine. We tested the set of reconstructed pathways against several requirements, such as thermodynamic feasibility and mass balance constraints, and discarded the pathways that were not biologically meaningful.[6] Next, we ranked the pruned pathways based on the several criteria, such as yield, number of known reaction steps, pathway length, etc. The steps of the employed workflow are discussed further (Figure 5).

Figure 5. Computational pipeline for discovery, evaluation and analysis of biosynthetic pathways.

Metabolic network generation

We applied the retrobiosynthesis algorithm of BNICE.ch[6] to generate a biosynthetic network that contains all theoretically possible compounds and reactions that are up to 5 reaction steps away from MEK. Starting from MEK and 26 cofactors appearing in the central carbon metabolism of living organisms

(Table S2 in the Supporting Information), we identified iteratively in a backward manner the reactions that lead to MEK along with its potential precursors.[46] The BNICE.ch network generation algorithm utilizes the expert-curated generalized enzyme reaction rules[18, 19, 47] for identifying all potential compounds and reactions that lead to the production of the target molecules. The most recent version of BNICE.ch includes 361*2 bidirectional generalized reaction rules capable of reconstructing more than 6'500 KEGG reactions.[21]

Note that for studies where we need to generate a metabolic network that involves only KEGG compounds, mining the ATLAS of Biochemistry[21] is a more efficient procedure than using BNICE.ch retrobiosynthesis algorithm. The ATLAS of Biochemistry is a repository that contains all KEGG reactions and over 130'000 novel enzymatic reactions between KEGG compounds.

Pathway reconstruction

We performed a graph-based search to reconstruct all possible pathways that connect the five target molecules with the set of 157 native *E. coli* metabolites (Table S3 in the Supporting Information)[32]. We reconstructed the exhaustive set of pathways up to the length of 4 reaction steps.

Note: If we were interested in pathways containing only KEGG reactions, we would perform a graph-based search over the network mined from the ATLAS of Biochemistry.

Pathway evaluation

It is crucial to identify and select, out of a vast number of generated pathways, the ones that satisfy physico-chemical constraints, such as mass balance and

thermodynamics, or the ones that have an economically viable production yield of the target compounds from a carbon source. Evaluation of pathways is context-dependent, and it is important to perform it in an exact host organism model and under the same physiological conditions as the ones that will be used in the experimental implementation.

Flux balance and thermodynamic-based flux balance analysis. We embedded the generated pathways one at the time in the latest genome-scale model of *E. coli*, iJO1366,[39] and we performed both Flux Balance Analysis (FBA)[48] and Thermodynamic-based Flux Analysis (TFA)[49-53] on the resulting models. In these analyses, we assumed that the only carbon source was glucose and we applied the following two sets of constraints on reaction directionalities:

(*C1*) We removed the preassigned reaction directionalities[54] from the iJO1366 model with the exception of ATP maintenance (ATPM), and we assumed that the reactions that involve CO_2 are operating in the decarboxylation direction. The lower bound on ATPM was set to 8.39 mmol/gDCW/hr. The remaining reactions were assumed to be bi-directional for FBA, whereas for TFA the directionality of these reactions was imposed by thermodynamics. The purpose of removing preassigned reaction directionalities was to explore the limitations that are imposed only by the physico-chemical properties of metabolic network.

(*C2*) This set of constraints contains the preassigned reaction directionalities from iJO1366 together with the constraints from *C1*.

Since FBA is less computationally expensive than TFA, we first performed FBA as a prescreening method to identify and discard the pathways: (i) that are not

satisfying the mass balance, e.g., pathways that need co-substrates not present in the model; and (ii) that have a yield from glucose to the target compounds lower than a pre-specified threshold. We then performed TFA on the reduced set of pathways to identify the pathways that are bio-energetically favorable and we computed their yields from glucose to 5 target compounds under thermodynamic constraints.

BridgIT analysis. We used BridgIT, our in-house developed computational tool, to find known reactions with associated genes in databases that were the most structurally similar to novel reactions appearing in the feasible pathways. BridgIT integrates the information about the structures of substrates and products of a reaction into reaction fingerprints.[55] These reaction fingerprints contain the information about chemical groups in substrates and products that were modified in the course of a reaction. BridgIT compares the reaction fingerprints of novel reactions to the ones of known reactions, and quantifies this comparison with the Tanimoto similarity score. The Tanimoto score of 1 signifies that two compared reactions had a high similarity, whereas the Tanimoto score values close to 0 signify that there was no similarity. We used this score to rank the reactions identified as similar to each of the novel reactions. The gene and protein sequences of the highest ranked reactions were proposed as candidates for either a direct experimental implementation or enzyme engineering.

Subnetwork reconstruction analysis

Once the most biologically feasible pathways were identified, we analyzed the parts of the metabolism that carry fluxes when the target compounds are

produced from glucose. We considered that the active parts of metabolism consisted of: (i) the core metabolic network (Figure 2, panel A), which included the central carbon pathways, such as glycolysis, pentose phosphate pathway, tricarboxylic cycle, electron transport chain; and (ii) the active non-core metabolic subnetworks (Figure 2, panel A), which contain of reactions that would carry fluxes when a target molecule is produced, but did not belong to the core metabolic network. We also defined the core precursors as metabolites that are connecting the core and the active non-core metabolic subnetworks (Figure 2, panel a).

We derived the core metabolic network from the genome-scale reconstruction iJO1366[39] using the redGEM algorithm[56], and we then used the lumpGEM[57] algorithm to identify active non-core subnetworks, and to compute their lumped reactions. The analysis of lumped reactions allowed us to identify core precursors of the target chemicals. We then performed clustering to uncover common enzymes, core precursors and intermediate metabolites of the non-core subnetworks leading to the production of the target chemicals.

Identification and lumping of active non-core subnetworks. The lumpGEM algorithm was applied to identify the comprehensive set of smallest metabolic subnetworks that were stoichiometrically balanced and capable of synthesizing a target compound from a defined set of core metabolites. The set of core metabolites belongs to the core metabolic network, and it includes also cofactors, small metabolites, and inorganic metabolites (Table S7 in the Supporting Information). Then, for each target compound and for each identified subnetwork, we used lumpGEM to collapse this subnetwork and generate a

corresponding lumped reaction. Within this process, we also identified the cost of core metabolites for the biosynthesis of these target compounds.

Clustering of subnetworks. To better understand the chemistry that leads towards the target compounds, we performed two types of clustering on the identified subnetworks:

- Clustering based on the structural similarity between the core precursors and byproducts of the lumped reactions. For each lumped reaction, we removed all non-carbon compounds, such as H_2, O_2, and phosphate, and the cofactor pairs, such as ATP and ADP, NAD^+ and NADH, $NADP^+$ and NADPH, flavodoxin oxidized and reduced, thioredoxin oxidized and reduced, ubiquinone and ubiquinol. This way, we created a set of substrates (core precursors) and byproducts of interest for each lumped reaction. We then used the *msim* algorithm from the RxnSim tool[58] to compare the lumped reactions based on individual similarities of their core precursors and byproducts. We finally used the obtained similarity scores to perform the clustering.
- Clustering based on the structural similarity between reactions that constitute the non-core subnetworks. We used BridgIT to compute structural fingerprints of reactions that constitute the non-core subnetworks, and we then performed a pairwise comparison of the non-core subnetworks as follows.

  For a given pair of non-core subnetworks, we carried out a pairwise comparison of their reactions. As a comparison metric we used the Tanimoto distance of the reaction fingerprints.[59] Based on this comparison, we found

the pair of the most similar reactions in two subnetworks and we stored the corresponding distance score. We then removed this pair of reactions from comparison, and we found the next pair of the most similar reactions, we stored their distance score, and we continued with this procedure until we found all pairs of reactions in two subnetworks. Whenever the number of reactions in two subnetworks was unequal, we ignored the unmatched reactions. The distance score between two compared subnetworks was formed as the sum of the distance scores of compared pairs of reactions. This procedure was repeated for all pairs of subnetworks.

We then used the computed distance scores to perform the subnetworks clustering.

Ranking and visualization of *in silico* pathways

In this step, we identified the pathways that were most likely to produce the target molecules. For scoring and ranking the biologically meaningful pathways we used several criteria: (i) maximum yield from glucose to the target molecules; (ii) minimal number of novel reactions, i.e., enzymes to be engineered; (iii) minimal number of reaction steps in the production pathway; and (iv) highest similarity scores from BridgIT.

Conclusions

In this work, we used BNICE.ch to reconstruct, evaluate and analyze more than 3.6 million biosynthetic pathways from the central carbon metabolites of *E. coli* toward five precursors of Methyl Ethyl Ketone (MEK), a 2nd generation biofuel

candidate. Our evaluation and analysis showed that more than 18'000 of these pathways are biologically feasible. We provided gene and protein sequences of the structurally most similar KEGG reactions to the novel reactions in the feasible pathways, which is valuable information for their experimental realization. Implementation of the discovered pathways in *E. coli* will allow the sustainable and efficient production of five precursors of MEK, 3-oxopentanoate, but-3-en-2-one, but-1-en-2-olate, butylamine, and 2-hydroxy-2-methylbutanenitrile, which can also be used as precursors for the production of other valuable chemicals.[36-38]

The pathway analysis methods developed and used in this work offer a systematic way for classifying and evaluating alternative ways for the production of target molecules. They also provide a better understanding of the underlying chemistry and can be used to guide the design of novel biosynthetic pathways for a wide range of biochemicals and for their implementation into host organisms.
The present study shows the potential of computational retrobiosynthetic tools for discovery and design of novel synthetic pathways, and their relevance for future developments in the area of metabolic engineering and synthetic biology.

Acknowledgments

M.T. was supported by the Ecole Polytechnique Fédérale de Lausanne (EPFL) and the ERASYNBIO1-016 SynPath project funded through ERASynBio Initiative for the robust development of Synthetic Biology. N.H. and M.A were supported through the RTD grant MicroScapesX, no. 2013/158, within SystemX, the Swiss Initiative for System Biology evaluated by the Swiss National Science

Foundation. L.M. and V.H. were supported by the Ecole Polytechnique Fédérale de Lausanne (EPFL).

Author information

* Corresponding author:

Vassily Hatzimanikatis,

Laboratory of Computational Systems Biotechnology (LCSB),

École Polytechnique Fédérale de Lausanne (EPFL),

CH-1015 Lausanne, Switzerland

Email: vassily.hatzimanikatis@epfl.ch

Phone: + 41 (0)21 693 98 70 Fax: +41 (0)21 693 98 75

Supporting information

Tables S1-S7 (XLSX)

S1: List of compounds one step away from MEK.

S2: List of starting compounds used for the retrobiosynthesis of BNICE.ch

S3: List of 157 starting compounds.

S4: Yield histograms for 5 MEK precursors obtained with C1 constraints.

S5: Yield histograms for 5 MEK precursors obtained with C2 constraints.

S6: List of 185 alternative pathways from AcCoA and PpCoA to 3-oxopentanoate.

S7: List of metabolites in the core metabolic network.

Figures 1-2 (PDF)

F1: Three alternative ways to produce 3-oxopentanoate from acetate through 2 intermediate metabolites: 2-ethylmalate and 3-hydroxypentanoate.

F2: An example of three different pathways from acetate to 3-oxopentanoate which share the same lumped reaction.

Abbreviations

PPM2	Phosphopentomutase 2 (deoxyribose)	5.4.2.7
DRPA	Deoxyribose-phosphate aldolase	4.1.2.4
ALDD3y	Aldehyde dehydrogenase (propanal, NADP)	1.2.1.4
3OXCOAT	3-oxoadipyl-CoA thiolase	2.3.1.174
MMM	Methylmalonyl-CoA mutase	5.4.99.2
AMPN	AMP nucleosidase	3.2.2.4
NTD6	5'-nucleotidase (dAMP)	3.1.3.89
PUNP2	Purine-nucleoside phosphorylase	2.4.2.1
ASPK	Aspartate kinase	2.7.2.4
THRD_L	L-threonine deaminase	4.1.1.19
RNDR1	Ribonucleoside-diphosphate reductase (ADP)	1.17.4.1
DADK	Deoxyadenylate kinase	2.7.4.3
NTTP5	Nucleoside triphosphate pyrophosphorylase	3.6.1.19
RNTR1c2	Ribonucleoside-triphosphate reductase (ATP)	1.17.4.2
MGSA	Methylglyoxal synthase	4.2.3.3
ACACT1r	Acetyl-CoA C-acetyltransferase	2.3.1.9
FTHFLi	Formate-tetrahydrofolate ligase	6.3.4.3
MCITL2	Methylisocitrate lyase	4.1.3.30

References:

1. Fischer, C. R., Klein-Marcuschamer, D., and Stephanopoulos, G. (2008) Selection and optimization of microbial hosts for biofuels production, *Metabolic engineering 10*, 295-304.
2. Atsumi, S., Hanai, T., and Liao, J. C. (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels, *Nature 451*, 86-89.
3. Schirmer, A., Rude, M. a., Li, X., Popova, E., and del Cardayre, S. B. (2010) Microbial biosynthesis of alkanes, *Science (New York, N.Y.) 329*, 559-562.
4. Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B., and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass, *Nature 463*, 559-562.
5. Cho, A., Yun, H., Park, J. H., Lee, S. Y., and Park, S. (2010) Prediction of novel synthetic pathways for the production of desired chemicals, *Bmc Systems Biology 4*.
6. Hadadi, N., and Hatzimanikatis, V. (2015) Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways, *Current Opinion in Chemical Biology 28*, 99-104.
7. Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2010) Discovery and Analysis of Novel Metabolic Pathways for the Biosynthesis of Industrial Chemicals: 3-Hydroxypropanoate, *Biotechnology and Bioengineering 106*, 462-473.
8. 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 Research 38*, W138-W143.
9. Cho, A., Yun, H., Park, J. H., Lee, S. Y., and Park, S. (2010) Prediction of novel synthetic pathways for the production of desired chemicals, *BMC Systems Biology 4*, 35.
10. Hou, B. K., Ellis, L. B. M., and Wackett, L. P. (2004) Encoding microbial metabolic logic: predicting biodegradation, *Journal of Industrial Microbiology and Biotechnology 31*, 261-272.
11. Ellis, L. B. M., Gao, J., Fenner, K., and Wackett, L. P. (2008) The University of Minnesota pathway prediction system: predicting metabolic logic, *Nucleic Acids Research 36*, W427-W432.
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, *Metabolic Engineering 25*, 140-158.
13. 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.
14. Carbonell, P., Parutto, P., Herisson, J., Pandit, S. B., and Faulon, J. L. (2014) XTMS: pathway design in an eXTended metabolic space, *Nucleic Acids Research 42*, W389-W394.

15. Heath, A. P., Bennett, G. N., and Kavraki, L. E. (2010) Finding metabolic pathways using atom tracking, *Bioinformatics 26*, 1548-1555.
16. Dale, J. M., Popescu, L., and Karp, P. D. (2010) Machine learning methods for metabolic pathway prediction, *Bmc Bioinformatics 11*.
17. Prather, K. L. J., and Martin, C. H. (2008) De novo biosynthetic pathways: rational design of microbial chemical factories, *Current Opinion in Biotechnology 19*, 468-474.
18. 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.
19. Hatzimanikatis, V., Li, C. H., Ionita, J. A., and Broadbelt, L. J. (2004) Metabolic networks: enzyme function and metabolite structure, *Curr Opin Struc Biol 14*, 300-306.
20. Hadadi, N., Soh, K. C., Seijo, M., Zisaki, A., Guan, X. L., Wenk, M. R., and Hatzimanikatis, V. (2014) A computational framework for integration of lipidomics data into metabolic pathways, *Metabolic Engineering 23*, 1-8.
21. Hadadi, N., Hafner, J., Shajkofci, A., Zisaki, A., and Hatzimanikatis, V. (2016) ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies, *ACS Synthetic Biology*.
22. Soh, K. C., and Hatzimanikatis, V. (2010) Dreams of Metabolism, *Trends Biotechnol 28*, 501-508.
23. Brunk, E., Neri, M., Tavernelli, I., Hatzimanikatis, V., and Rothlisberger, U. (2012) Integrating computational methods to retrofit enzymes to synthetic pathways, *Biotechnology and Bioengineering 109*, 572-582.
24. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol, *Nature Chemical Biology 7*, 445-452.
25. Ghiaci, P., Norbeck, J., and Larsson, C. (2014) 2-Butanol and Butanone Production in *Saccharomyces cerevisiae* through Combination of a B12 Dependent Dehydratase and a Secondary Alcohol Dehydrogenase Using a TEV-Based Expression System, *PLoS ONE 9*, e102774.
26. Hoell, D., Mensing, T., Roggenbuck, R., Sakuth, M., Sperlich, E., Urban, T., Neier, W., and Strehlke, G. (2009) 2-Butanone, In *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA.
27. Hoppe, F., Burke, U., Thewes, M., Heufer, A., Kremer, F., and Pischinger, S. (2016) Tailor-Made Fuels from Biomass: Potentials of 2-butanone and 2-methylfuran in direct injection spark ignition engines, *Fuel 167*, 106-117.
28. Srirangan, K., Liu, X., Akawi, L., Bruder, M., Moo-young, M., and Chou, C. P. (2016) Engineering Escherichia coli for Microbial Production of Butanone, 2574-2584.
29. Yoneda, H., Tantillo, D. J., and Atsumi, S. (2014) Biological production of 2-butanone in Escherichia coli, *ChemSusChem 7*, 92-95.
30. Multer, A., McGraw, N., Hohn, K., and Vadlani, P. (2013) Production of Methyl Ethyl Ketone from Biomass Using a Hybrid Biochemical/Catalytic Approach, *Industrial & Engineering Chemistry Research 52*, 56-60.

31. Drabo, P., Tiso, T., Heyman, B., Sarikaya, E., Gaspar, P., Förster, J., Büchs, J., Blank, L. M., and Delidovich, I. (2017) Anionic Extraction for Efficient Recovery of Biobased 2,3-Butanediol—A Platform for Bulk and Fine Chemicals, *ChemSusChem 10*, 3252-3259.
32. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation, *Nucleic Acids Research 44*, D457-D462.
33. Kanehisa, M., and Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Research 28*, 27-30.
34. Engel, T. (2007) The structural- and bioassay database PubChem, *Nachr Chem 55*, 521-524.
35. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., and Bryant, S. H. (2016) PubChem Substance and Compound databases, *Nucleic Acids Research 44*, D1202-D1213.
36. Krumpfer, J. W., Giebel, E., Frank, E., Müller, A., Ackermann, L.-M., Tironi, C. N., Mourgas, G., Unold, J., Klapper, M., Buchmeiser, M. R., and Müllen, K. (2017) Poly(Methyl Vinyl Ketone) as a Potential Carbon Fiber Precursor, *Chemistry of Materials 29*, 780-788.
37. Siegel, H., and Eggersdorfer, M. (2000) Ketones, In *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA.
38. Eller, K., Henkes, E., Rossbacher, R., and Höke, H. (2000) Amines, Aliphatic, In *Ullmann's Encyclopedia of Industrial Chemistry*, Wiley-VCH Verlag GmbH & Co. KGaA.
39. Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., and Palsson, B. O. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011, *Molecular Systems Biology 7*.
40. Islam, M. A., Hadadi, N., Ataman, M., Hatzimanikatis, V., and Stephanopoulos, G. (2017) Exploring biochemical pathways for mono-ethylene glycol (MEG) synthesis from synthesis gas, *Metabolic Engineering 41*, 173-181.
41. Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990) *Physiology of the bacterial cell : a molecular approach*, Sinauer Associates, Sunderland, Mass.
42. Sudarsan, S., Dethlefsen, S., Blank, L. M., Siemann-Herzberg, M., and Schmid, A. (2014) The Functional Structure of Central Carbon Metabolism in Pseudomonas putida KT2440, *Applied and Environmental Microbiology 80*, 5292-5303.
43. Lilley, D. M. J., Clegg, R. M., Diekmann, S., Seeman, N. C., Von Kitzing, E., and Hagerman, P. J. (1995) A nomenclature of junctions and branchpoints in nucleic acids, *Nucleic Acids Research 23*, 3363-3364.
44. Chen, Y., Daviet, L., Schalk, M., Siewers, V., and Nielsen, J. (2013) Establishing a platform cell factory through engineering of yeast acetyl-CoA metabolism, *Metab Eng 15*.
45. Haller, T., Buckel, T., Rétey, J., and Gerlt, J. A. (2000) Discovering New Enzymes and Metabolic Pathways: Conversion of Succinate to Propionate by Escherichia coli, *Biochemistry 39*, 4622-4629.
46. Corey, E. J. (1991) The Logic of Chemical Synthesis - Multistep Synthesis of Complex Carbogenic Molecules, *Angew Chem Int Edit 30*, 455-465.

47. Li, C. H., Henry, C. S., Jankowski, M. D., Ionita, J. A., Hatzimanikatis, V., and Broadbelt, L. J. (2004) Computational discovery of biochemical routes to specialty chemicals, *Chemical Engineering Science 59*, 5051-5060.
48. Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010) What is flux balance analysis?, *Nature biotechnology 28*, 245-248.
49. Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007) Thermodynamics-based metabolic flux analysis, *Biophys J 92*, 1792-1805.
50. Henry, C. S., Jankowski, M. D., Broadbelt, L. J., and Hatzimanikatis, V. (2006) Genome-scale thermodynamic analysis of Escherichia coli metabolism, *Biophys J 90*, 1453-1461.
51. Ataman, M., and Hatzimanikatis, V. (2015) Heading in the right direction: thermodynamics-based network analysis and pathway engineering, *Curr Opin Biotechnol 36*, 176-182.
52. Soh, K. C., and Hatzimanikatis, V. (2010) Network thermodynamics in the post-genomic era, *Curr Opin Microbiol 13*, 350-357.
53. Soh, K. S., and Hatzimanikatis, V. (2014) Constraining the flux space using thermodynamics and integration of metabolomics data, *Methods in Molecular Biology 1191*, 49-63.
54. Frainay, C., and Jourdan, F. (2017) Computational methods to identify metabolic sub-networks based on metabolomic profiles, *Briefings in Bioinformatics 18*, 43-56.
55. James, C. A., and Weininger, D. Daylight Theory Manual, *Daylight Chemical Information Systems, Inc.: Irvine, CA*.
56. Ataman, M., Gardiol, D. H. F., Fengos, G., and Hatzimanikatis, V. (2017) redGEM: Systematic Reduction and Analysis of Genome-scale Metabolic Reconstructions for Development of Consistent Core Metabolic Models, *Plos Computational Biology*.
57. Ataman, M., and Hatzimanikatis, V. (2017) lumpGEM: Systematic Generation of Subnetworks and Elementally Balanced Lumped Reactions for the Biosynthesis of Target Metabolites, *Plos Computational Biology*.
58. Giri, V., Sivakumar, T. V., Cho, K. M., Kim, T. Y., and Bhaduri, A. (2015) RxnSim: a tool to compare biochemical reactions, *Bioinformatics 31*, 3712-3714.
59. Bajusz, D., Rácz, A., and Héberger, K. (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?, *Journal of Cheminformatics 7*, 20.

Figures

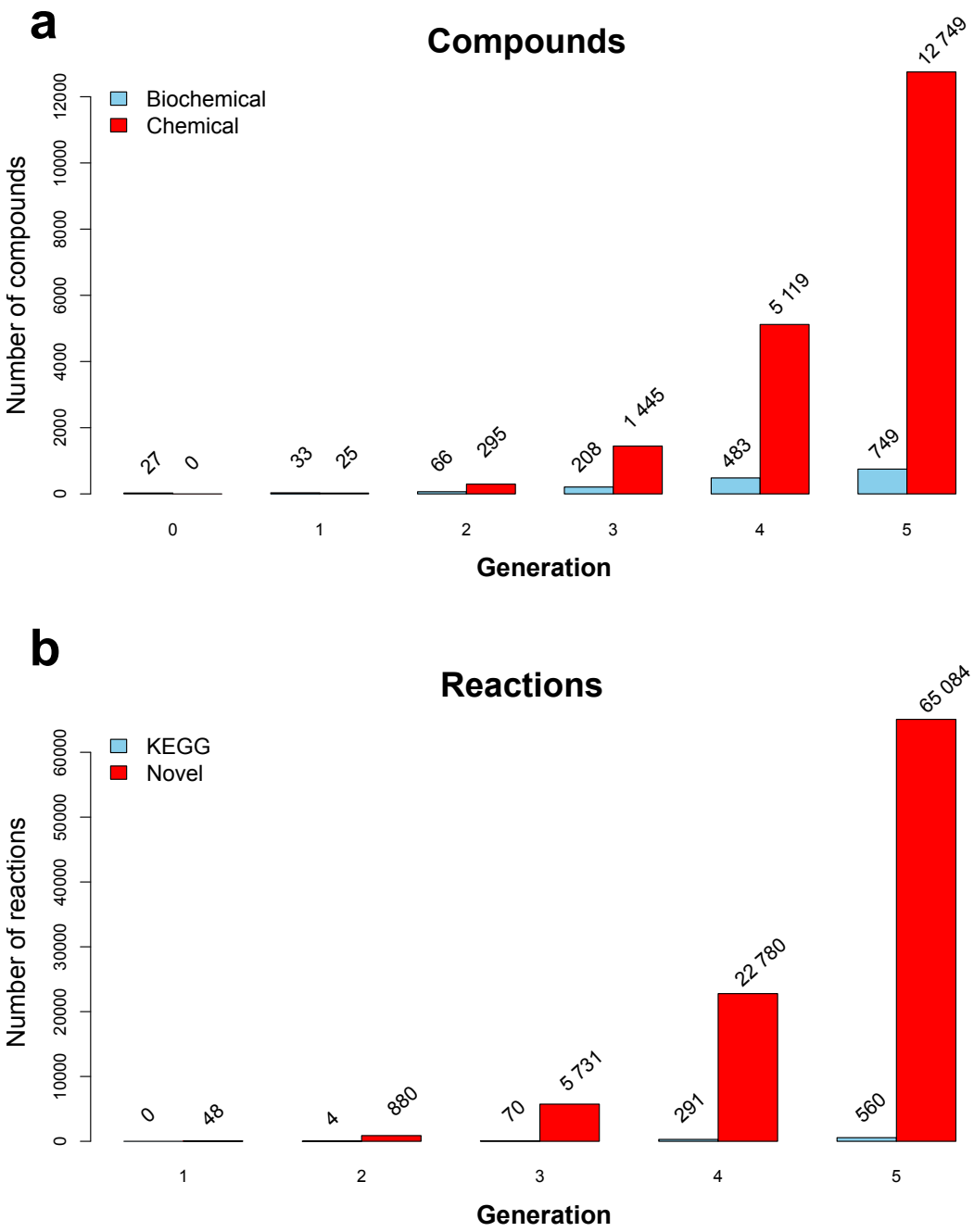


Figure 1

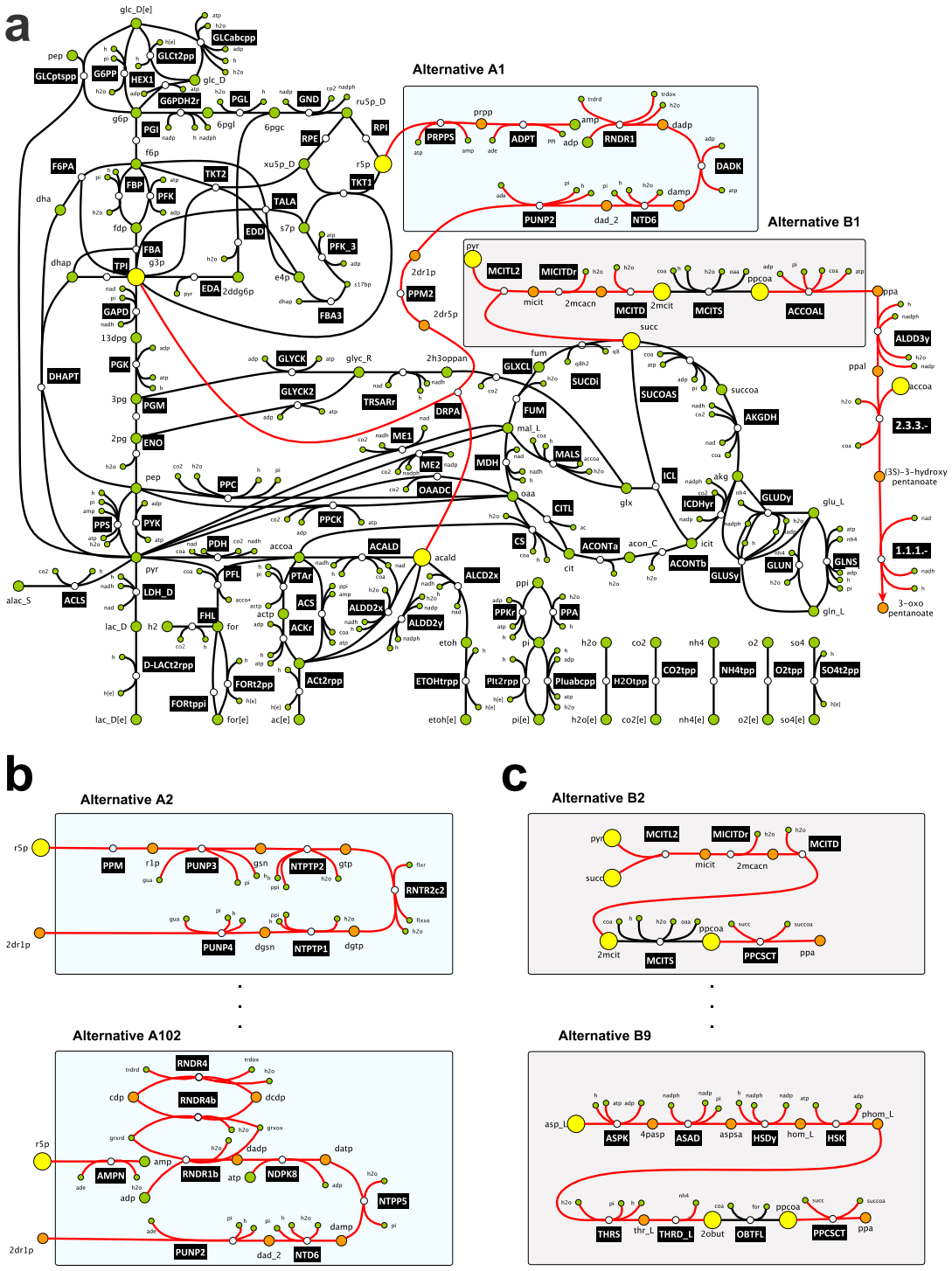

a
Alternative A1
Alternative B1
b
Alternative A2
Alternative A102
c
Alternative B2
Alternative B9

Figure 2

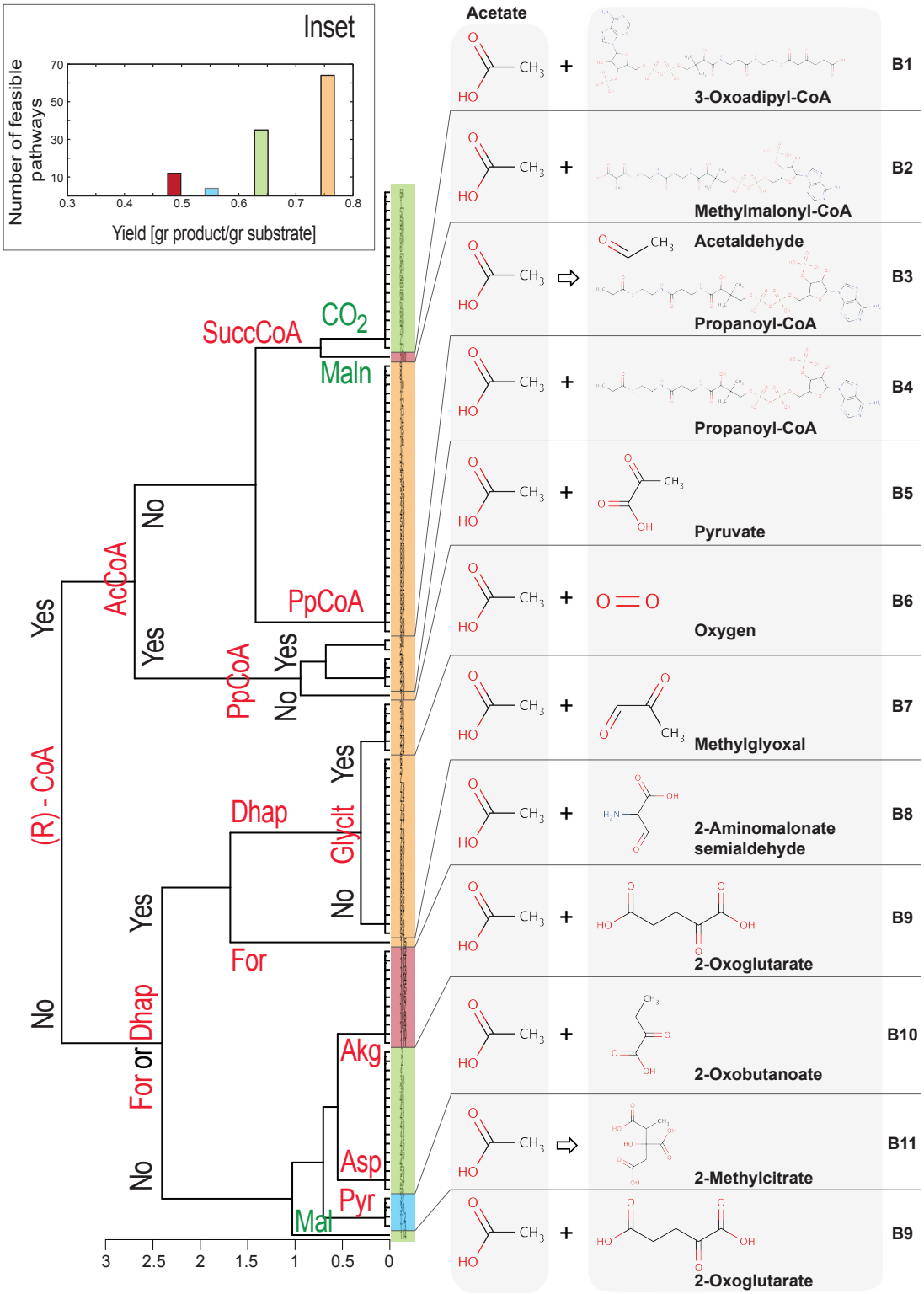

Inset
Number of feasible pathways
Yield [gr product/gr substrate]
Acetate
3-Oxoadipyl-CoA
Methylmalonyl-CoA
Acetaldehyde
Propanoyl-CoA
Propanoyl-CoA
Pyruvate
Oxygen
Methylglyoxal
2-Aminomalonate semialdehyde
2-Oxoglutarate
2-Oxobutanoate
2-Methylcitrate
2-Oxoglutarate
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B9
SuccCoA
CO2
Maln
AcCoA
PpCoA
PpCoA
Dhap
Glyclt
For
Akg
Asp
Pyr
Mal
(R) - CoA
Yes
No
Yes
No
Yes
No
Yes
No
Yes
No
Yes
No
3
2.5
2
1.5
1
0.5
0

Figure 3

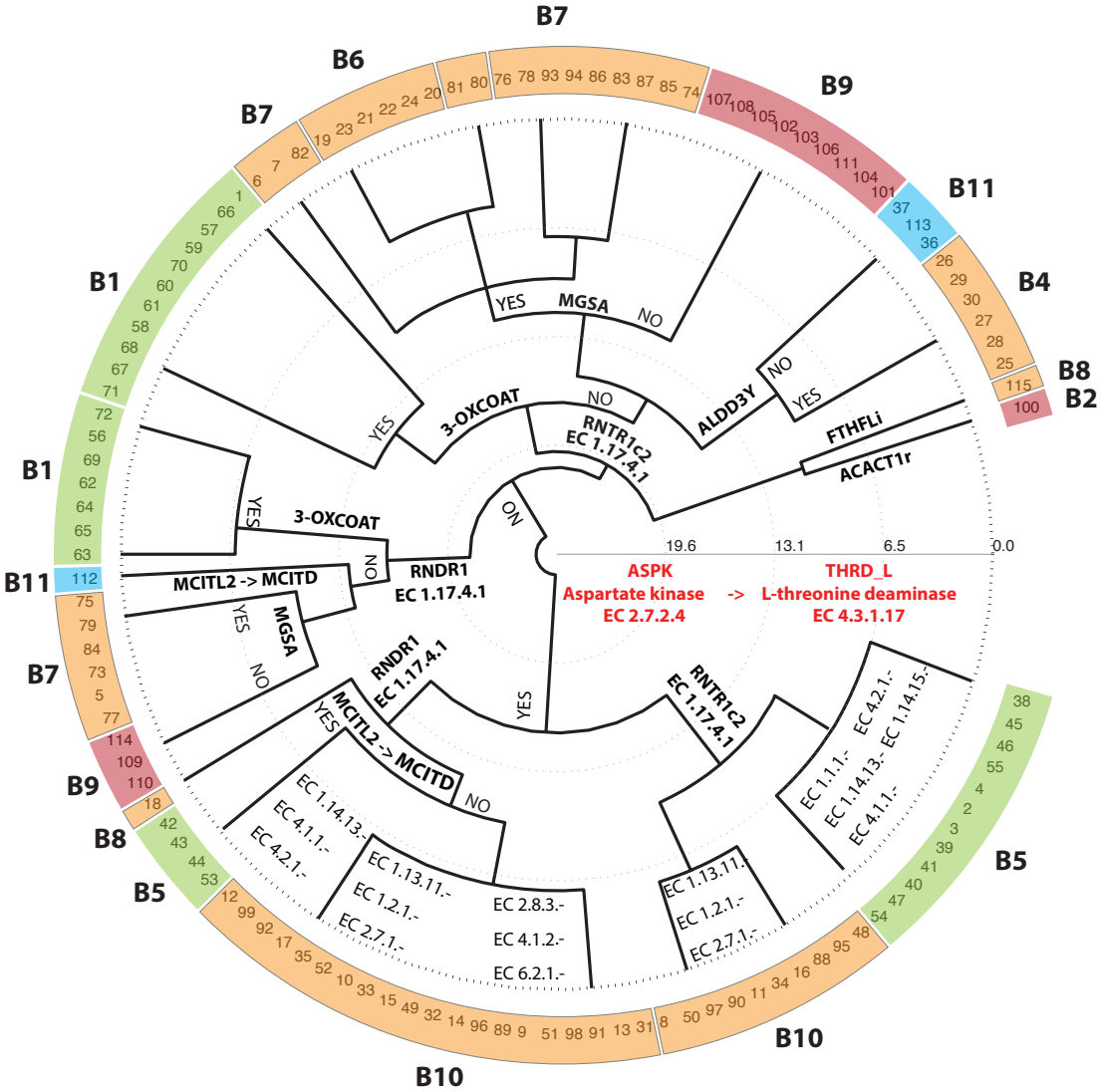

B7
B6
B7
B9
B11
B4
B8
B2
B1
B1
B11
B7
B9
B8
B5
B5
B10
B10
MGSA
YES
NO
3-OXCOAT
YES
NO
RNTR1c2
EC 1.17.4.1
ALDD3Y
NO
YES
FTHFLi
ACACT1r
3-OXCOAT
YES
NO
RNDR1
EC 1.17.4.1
MCITL2 -> MCITD
MGSA
YES
NO
RNDR1
EC 1.17.4.1
MCITL2 -> MCITD
YES
NO
YES
RNTR1c2
EC 1.17.4.1
19.6
13.1
6.5
0.0
ASPK
Aspartate kinase
EC 2.7.2.4
->
THRD_L
L-threonine deaminase
EC 4.3.1.17
EC 1.14.13.-
EC 4.1.1.-
EC 4.2.1.-
EC 1.13.11.-
EC 1.2.1.-
EC 2.7.1.-
EC 2.8.3.-
EC 4.1.2.-
EC 6.2.1.-
EC 1.13.11.
EC 1.2.1.-
EC 2.7.1.-
EC 1.1.1.-
EC 1.14.13.-
EC 4.1.1.-
EC 4.2.1.-
EC 1.14.15.-

Figure 4

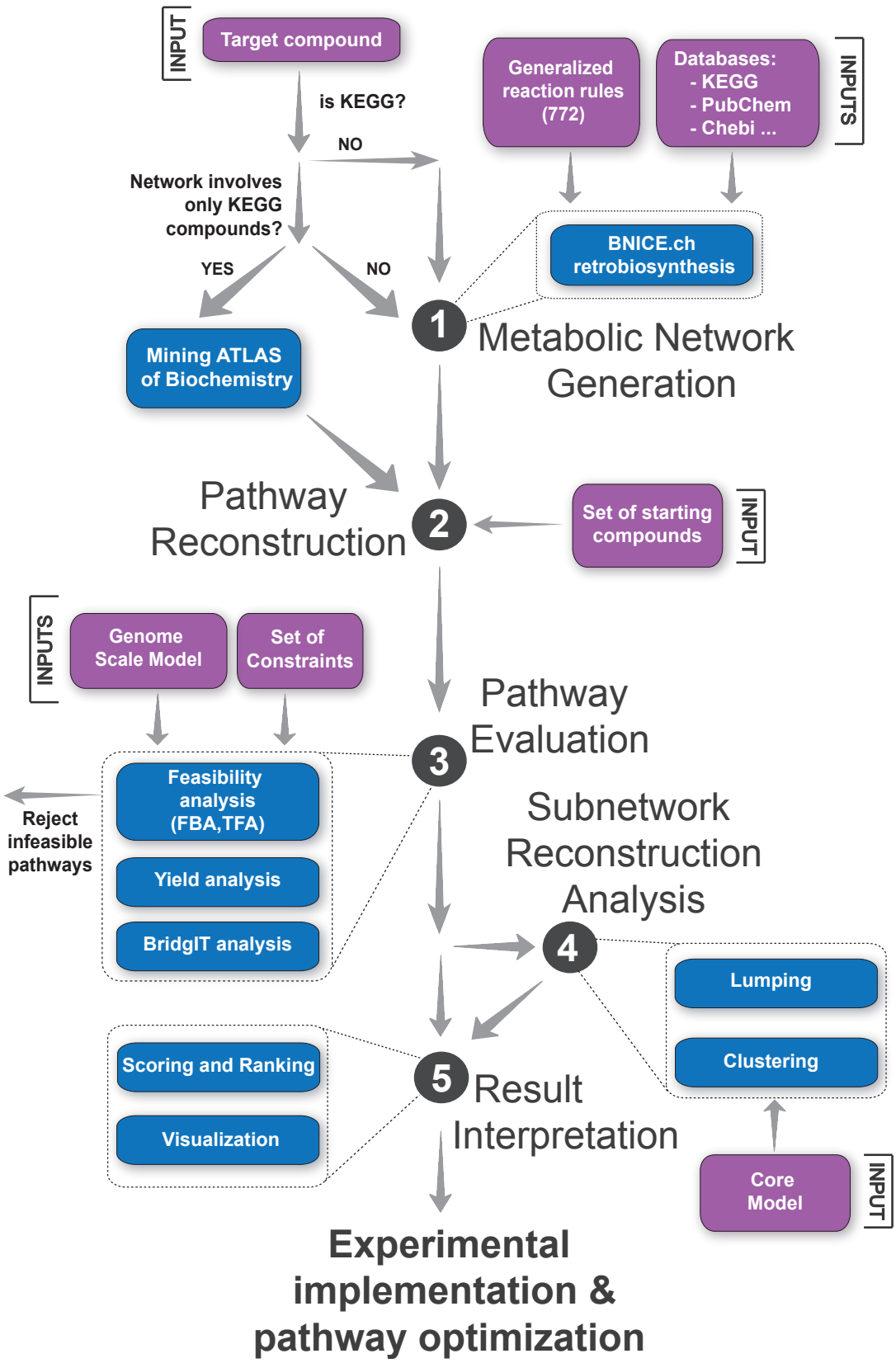

INPUT
Target compound
is KEGG?
NO
Network involves only KEGG compounds?
YES
NO
Generalized reaction rules (772)
Databases:
- KEGG
- PubChem
- Chebi ...
INPUTS
BNICE.ch retrobiosynthesis
1
Metabolic Network Generation
Mining ATLAS of Biochemistry
Pathway Reconstruction
2
Set of starting compounds
INPUT
INPUTS
Genome Scale Model
Set of Constraints
Pathway Evaluation
3
Feasibility analysis (FBA,TFA)
Reject infeasible pathways
Yield analysis
BridgIT analysis
Subnetwork Reconstruction Analysis
4
Lumping
Clustering
Scoring and Ranking
Visualization
5
Result Interpretation
Core Model
INPUT
Experimental implementation & pathway optimization

Figure 5