Towards synthetic biology strategies for adipic acid production – an *in-silico* tool for combined thermodynamics and stoichiometric analysis of metabolic networks

Nils J. H. Averesch[#a,b*], Verónica S. Martínez[#c,d*], Lars K. Nielsen[c,e], Jens O. Krömer[a,f]

[a] Centre for Microbial Electrochemical Systems (CEMES), Advanced Water Management Centre (AWMC), The University of Queensland, Brisbane, Australia

[b] Universities Space Research Association at NASA Ames Research Center, Moffett Field, California

[c] Systems and Synthetic Biology Group, Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, Australia

[d] ARC Training Centre for Biopharmaceutical Innovation (CBI), Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, Australia

[e] DTU BIOSUSTAIN, Novo Nordisk Foundation Center for Biosustainability, Danmarks Tekniske Universitet, Kemitorvet, Denmark

[f] Department for Solar Materials, Helmholtz Centre of Environmental Research - UFZ, Leipzig, Germany

[*] corresponding authors

[#] contributed equally

Email addresses:

NJHA: nils.averesch@uq.net.au, VSM: v.salazar@uq.edu.au, LKN: lars.nielsen@uq.edu.au, JOK: jens.kroemer@ufz.de

Abstract

Adipic acid, a nylon-6,6 precursor, has recently gained popularity in synthetic biology. Here, 16 different production routes to adipic acid were evaluated using a novel tool for network-embedded thermodynamic analysis of elementary flux modes. The tool distinguishes between thermodynamically feasible and infeasible modes under determined metabolite concentrations, allowing the thermodynamic feasibility of theoretical yields to be assessed. Further, patterns that always caused infeasible flux distributions were identified, which will aid the development of tailored strain design.

A review of cellular efflux mechanisms revealed that significant accumulation of extracellular product is only possible if coupled with ATP hydrolysis. A stoichiometric analysis demonstrated that the maximum theoretical product carbon yield heavily depends on the metabolic route, ranging from 32% to 99% on glucose and/or palmitate in *E. coli* and *S. cerevisiae* metabolic models. Equally important, metabolite concentrations appeared to be thermodynamically restricted in several pathways. Consequently, the number of thermodynamically feasible flux distributions was reduced, in some cases even rendering whole pathways infeasible, highlighting the importance of pathway choice. Only routes based on the shikimate pathway were thermodynamically favourable over a large concentration and pH range. Low pH capability of *S. cerevisiae* shifted the thermodynamic equilibrium of some pathways towards product formation. One identified *infeasible-pattern* revealed that the reversibility of the mitochondrial malate dehydrogenase contradicted current state of knowledge, which imposes a major restriction on the metabolism of *S. cerevisiae*. Finally, the evaluation of industrially relevant constraints revealed that two shikimate pathway based routes in *E. coli* were most robust.

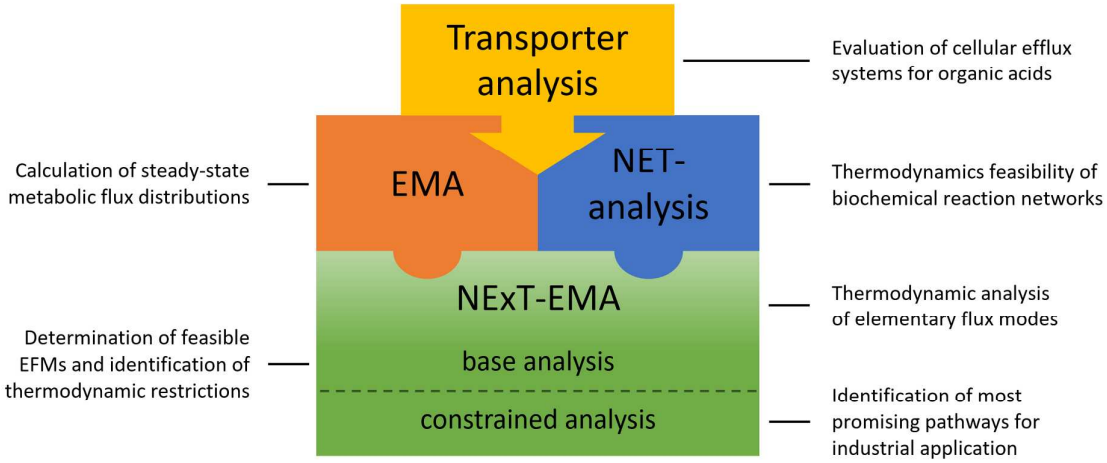


Table of Contents Graphic: Workflow of the analysis depicting how the respective parts are tied into the whole study.

Keywords

adipic acid; muconic acid; glucaric acid; elementary flux mode analysis; thermodynamics-based network analysis; organic acid transport mechanisms

Abbreviations

2,3-Ddha-CoA, 2,3-didehydroadipyl-CoA; Ac-CoA, acetyl-CoA; c_{ext}, external metabolite concentration; c_{int}, internal metabolite concentration; CoA, Coenzyme A; DHS, 3-dehydroshikimate; E4P, erythrose-4-phosphate; EFM, elementary flux mode; EMA, elementary mode analysis; FA, fatty acid; FAS, fatty acid synthesis; FVA, flux variability analysis; G6P, glucose-6-phosphate; GLC, glucose; GLN, glutamine, GLU, glutamate; MDH, malate dehydrogenase; NET-analysis, network-embedded thermodynamic analysis; OAA, oxaloacetate; 2-OXO, 2-oxoglutarate; P, phosphate; PEP, phosphoenol-pyruvate; $pH_{ext.}$, extracellular pH; $pH_{int.}$, intracellular pH; PYR, pyruvate; SUCC, succinate; SUCC-CoA, succinyl coenzyme A; TCA Cycle, tricarboxylic acid cycle; Y_{max}, maximum theoretical carbon yield; $\Delta_f G$, Gibbs free energy of formation; $\Delta_r G$, Gibbs free energy of reaction; $\Delta_r G'^{0}$, transformed standard Gibbs free energy of reaction

Adipic acid, or hexanedioic acid, is an essential precursor in the chemical industry. Its production is approximately 2.6 million tonnes per year [1, 2] and was valued at USD 4.55 billion in 2013 [3]. With an expected annual growth of 3.7 – 5%, the market is expected to reach USD 7.5 billion in 2019 [3, 4]. More than 65% of the adipic acid is high purity and used for nylon-6,6 production, while the rest is used for polyurethanes and a wide range of other applications [5-9]. Adipic acid is mainly produced through benzene reduction to cyclohexane followed by a two-step oxidation and reaction with nitric acid [10, 11]. While this method achieves very high yields, considerable amounts of NO_x, especially N_2O are generated. Consequently, adipic acid production accounts for almost 10% of the total global anthropogenic N_2O emissions per year, significantly contributing to global warming and depletion of the ozone layer [12-14].

Biotechnological production of adipic acid offers a promising alternative to chemical synthesis. Three major approaches have been reported: (i) direct production of adipic acid, (ii) production of the unsaturated adipic acid precursor *cis,cis*-muconic acid and (iii) production of D-glucaric acid. Muconic and glucaric acid can be converted to adipic acid via chemo-catalytic hydrogenation [15]; for *cis,cis*-muconic acid an efficient conversion yield of 97% [$mol_{adipate}/mol_{muconate}$] has been reported [16, 17]. After earlier works in the mid 1990's [18, 19], adipic acid has returned to the scientific spotlight [2, 8, 15, 20-24] and the metabolic engineering community has re-focussed on this compound [5, 16, 25-38]. The filing of patents by several companies (Genomatica, Verdezyne and other assignees) also shows a strong commercial interest to develop a biological replacement for fossil fuel derived adipic acid [39-43]. Considering the extensive work that has been performed on this topic, it is not surprising that there is a wide variety of proclaimed metabolic routes (figure 1).

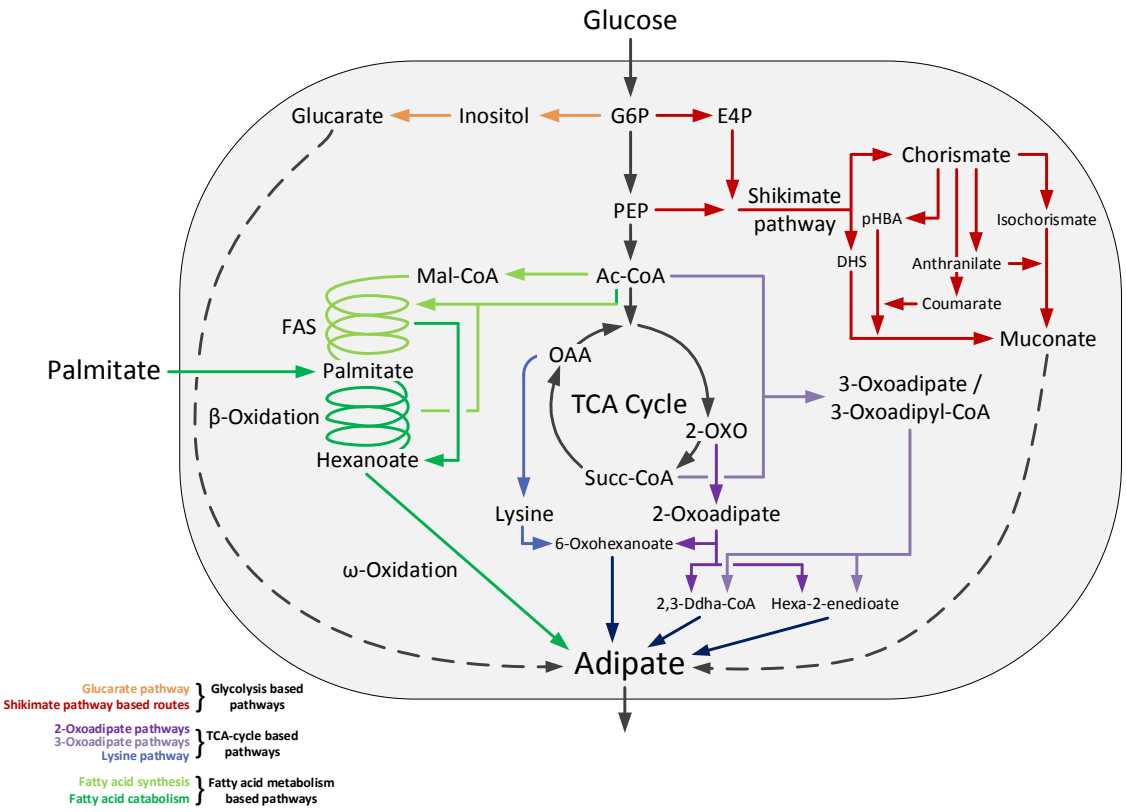


Figure 1: Overview of metabolic routes to adipic acid and its precursors, muconic and glucaric acid. Each major pathway has several sub-routes that utilize alternative biochemical conversions, which can significantly influence titer, yield and productivity. Biochemical reduction of muconic and glucaric acid to adipic acid, has not yet been accomplished and is therefore indicated as hypothetical with dashed lines.

Biotechnological production must outperform existing chemical synthesis in order to warrant additional investments. The cost advantage in comparison to petrochemical-derived adipic acid has been predicted to reach 20 – 30% [44]. Clearly, this cost advantage will strongly depend on the substrate(s) and pathway utilized to produce adipic acid, or its bio-precursors. The wealth of possibilities (figure 1) creates the challenge of finding the most promising route, a problem that has so far been ignored for adipic acid [45].

In general, optimization of a fermentation process requires yields, rates and titers to be maximized [46]. One approach to assess the yields of a metabolic network is elementary flux mode (EFM) analysis (EMA) [47]. EMA generates all elementary (i.e., non-decomposable)

feasible steady state flux distributions of a metabolic network in a single analysis and enables the assessment of maximum product yields as well as the determination of the ideal flux distribution in a network during optimal production. The approach has been applied to the development of strain construction strategies for a range of fermentation products, such as muconate, glutamate, methionine, succinate, isobutanol and other compounds [48-54].

Productivity or rates in a process are controlled by capacity- and kinetics-based regulation. While capacity is readily manipulated using gene knock-out / knock-in and over-expression, a reaction only occurs in the thermodynamically feasible direction, i.e. where the Gibbs free energy of reaction (Δ_rG) is negative [55-58]. The range of Δ_rG is estimated from the reactants' standard Gibbs free energies of formation (Δ_fG^0) and metabolite concentrations. In metabolism, individual irreversible reactions may dictate the range of feasible metabolite concentrations and therefore the direction of otherwise reversible reactions. Thus, thermodynamic constraints can propagate widely throughout a metabolic model. Many approaches exist for thermodynamic-based analysis of metabolic networks [59], one of these is network-embedded thermodynamic analysis (NET-analysis). NET-analysis determines thermodynamic feasibility of a full metabolic network [60] and can assess whether metabolite concentrations are feasible with respect to the reaction directionalities. In reverse, NET-analysis can also be used to assign reaction directionalities based on physiological ranges of metabolite concentrations and thermodynamic constraints [61]. This enables the prediction of a given pathway's thermodynamic feasibility based on metabolite concentration ranges. In the present context it is to be expected, that many stoichiometrically highly efficient routes are in fact thermodynamically infeasible given physiologically meaningful metabolite concentrations [62].

In the current study, we use NET-analysis to determine if individual EFMs are thermodynamically feasible. Thermodynamic analysis of EFMs was initially introduced in

combination with a genetic algorithm used to design *E. coli* metabolic networks optimized for product and biomass formation [62]. Full NET-analysis of EFMs was first applied in combination with flux variability analysis (FVA) to *S. cerevisiae* metabolism and revealed combinations of reactions that could not operate under specific conditions [63]. *tEFMA*, a tool that integrates evaluation of thermodynamic feasibility of EFMs during enumeration of these, was developed to facilitate EMA of large scale metabolic networks [64]. However, the tool was only demonstrated with a simple *E. coli* network, and while it allows the determination of patterns, which cause infeasibility, it is not able to quantify the effect of these patterns in the total number of infeasible EFMs. The same limitations apply to another recently published approach [65].

In the current study, we developed *NExT-EMA*, a tool that channels EFMs into NET-analysis. *NExT-EMA* was designed with ease of use in mind while not sacrificing broad applicability. It requires considerably less input than previous approaches (no FVA data needed) and is therefore readily applied and transferred to various metabolic networks. Further, it is based on the most recent and advanced thermodynamic data [66] and tool (*NExT*) for thermodynamic analysis of metabolic networks [67]. *NExT* is the latest implementation of NET-analysis and was previously demonstrated to handle large eukaryotic metabolic reconstructions including the human metabolic network [61]. *NExT-EMA* enables the analysis of fully compartmentalized complex eukaryotic networks, with an emphasis on rapid enumeration and minimization of computational costs (cf. section 2.5). Moreover, we developed an algorithm to determine the unique flux patterns (certain combinations of reaction directionalities) which cause infeasibility in EFMs. This novel capability is crucial to avoid unfavourable strain construction strategies.

Another factor that impacts product titer, yield and rate is the cellular mechanism of product export [68]. While uncharged molecules can generally diffuse through bio-membranes without

resistance, charged compounds like organic acids, are restricted by concentration and pH gradient [69]. Therefore, charged compounds often rely on ATP hydrolysis for export [70]. Here we evaluated four different mechanisms for the export of adipic acid, their particular stoichiometry, thermodynamics and the resulting significance for the export of carboxylic acids, which has so far mostly been neglected.

NExT-EMA was used to analyse *E. coli* and *S. cerevisiae* metabolic networks, comparing the yields and thermodynamic feasibility of all currently known routes to adipic acid and its biological precursors *cis,cis*-muconic acid and glucaric acid from glucose and/or fatty acids (figure 1). A "real-life scenario" considering industrially relevant conditions found increased thermodynamic constraints severely impact pathway feasibility. Our finding highlights the importance of pathway and organism choice to maximize the potential of a bio-based process, leading the metabolic engineering community towards highly efficient biotechnical production of adipic acid.

1. Results and Discussion

Thermodynamics can explain equilibria of chemical reactions and their spontaneous changes in a biological context. Applying these principles in synthetic biology, through comprehensive stoichiometric and thermodynamic analyses of an organism's biochemistry, can prevent the pursuit of physicochemically infeasible strain designs. Here we provide an implementation that enables a holistic analysis of cellular metabolism and demonstrate it on adipic acid production. First, a detailed evaluation of alternative product export mechanisms was used to determine the most efficient efflux system for organic acids in each organism. Elementary mode analysis of each pathway to adipic acid production revealed fundamental differences in the maximum possible carbon yield between the available biochemical options. Then, estimation of the reaction Gibbs free energies across a range of pH characterized the thermodynamic restrictions affecting the pathways. Subsequent thermodynamic analysis of EFMs excluded thermodynamically infeasible modes resulting in a higher fidelity determination of the theoretical yields. Finally, *NExT-EMA* was refined adding constraints that reflect the need for industrially relevant rates, yields and titers, thus exposing the most robust pathways.

1.1 Comparison of different product export scenarios

The efficiency and mechanism of the product export from the cells greatly affect the thermodynamic equilibria of the reactions in the product pathway and the achievable product titers in a process [68]. This is particularly important for pathways that have reactions with transformed standard Δ_rG ($\Delta_rG'^0$) close to zero, as limitations in export can lead to accumulation of intermediates and thus cause a reaction to stagnate because of thermodynamic infeasibility. Therefore, we evaluated the impact of different efflux systems for adipic acid in the cell envelope in detail.

Passive transport out of the cell via diffusion over the cell membrane is theoretically possible, as long as the driving force (the intracellular concentration of product) is high enough. Making use of (Eq. (2), section 2.1) to estimate adipic acid diffusion out of the cell, it becomes clear that c_{ext} will always be magnitudes smaller than c_{int} as long as pH_{int} is higher than pH_{ext} (in most cases given, due to the obligatory proton gradient across the membrane). With respective values for pH_{int} of 7.2 [71] and 7.6 [72] for *S. cerevisiae* and *E. coli* at pH_{ext} levels of 6.7 and 7.0, respectively, and assuming a physiologically very high intracellular concentration of adipic acid of 100 mM, a maximum extracellular concentration of 10.3 mM ($\approx 1.5\ g\times L^{-1}$) for *S. cerevisiae* and 6.4 mM ($\approx 0.9\ g\times L^{-1}$) for *E. coli* may be achieved by passive diffusion (supplementary file 1). When multiplying $10^{n\times\Delta pH}$ through on the right-hand side of Eq. (2) it becomes apparent that it is equivalent to Eq. (3). Therefore, transport of the neutral species by passive diffusion or transport of the dissociated form neutralized by proton symport are equivalent from an energetic point of view and the same results can be achieved when assuming a proton symporter. In contrast, transport of the fully dissociated acid is greatly favoured as it relies on actual energy usage. The ratio of internal to external adipic acid concentration is 3.9×10^{-4} and 1.5×10^{-14} for *E. coli* and *S. cerevisiae*, respectively (figure 2). Similarly, an active ABC-like transport system (Eq. (5), section 2.1) can achieve extracellular concentrations at the solubility limit (approx. 212 mM at 30°C and 335 mM at 37°C) with only minimal intracellular concentrations (c_{int} = 0.08 µM for *S. cerevisiae* and c_{int} = 0.13 µM for *E. coli*).

These observations are even more remarkable considering that in bacteria, as for example *E. coli*, (di)acids are believed to be mainly transported across the membrane (in both directions) by means of proton symporters [73, 74]. ABC transporters exist in *E. coli*, but are mainly specific for compounds like amino acids, sugars or vitamins [75]. In *S. cerevisiae* however, proton symporters are known to mostly mediate the influx of carboxylic acids, while efflux relies on

ATP-binding cassettes transporters [76], which also makes much more sense for the organism from a thermodynamic point of view [77]. This means that *E. coli* will almost certainly use an ATPase to balance dissociated acid transport, as it features a sufficiently high driving force and the expense is only 0.3 ATP per charge. In yeast this strategy is inefficient as the P-type ATPase uses 1 ATP per charge, i.e. 2 ATP per adipic acid transported, which is excessive [77]: 2 ATP per diacid ensure export against any gradient but severely impact the energy budget of the cell, therefore diminishing the product yield. An ABC transporter is adequate to ensure transport against any gradient, and ABC transporters of the Pleiotropic Drug Resistance (PDR) subfamily are known to be promiscuous carboxylate efflux pumps [76]. Therefore, transport of fully dissociated acid was assumed for *E. coli*, while an ABC transporter was assumed in *S. cerevisiae* for the subsequent metabolic network studies.

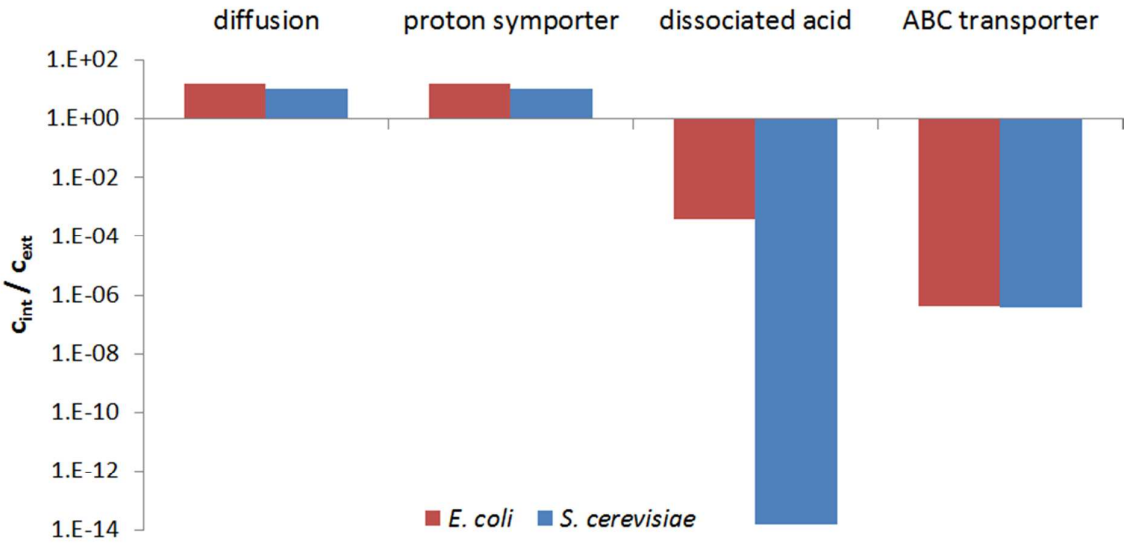


Figure 2: Logarithmic plot of internal and external adipic acid concentration ratios of the different studied transport mechanisms for *E. coli* (red) and *S. cerevisiae* (blue). The four different scenarios are passive diffusion of neutralized acid, proton anionic symporter, transport of dissociated acid with ATP usage, and ABC-transporter operating at equilibrium.

To date, product export has not been characterized in any *in vivo* study focusing on production of muconic, glucaric or adipic acid, but given the thermodynamic constraints that many pathways face in terms of feasible intracellular concentration ranges (cf. section 1.3 and following) this seems to be one of the most relevant aspects (besides toxicity – in fact, the differences in efflux may explain the much higher toxic effects of adipic acid on bacteria, compared to fungi [45]).

Many of the currently published titers of adipic and muconic acid are within the feasible range of passive diffusion, especially those of muconic acid production in *S. cerevisiae* [27, 28]. Also in *E. coli* many pathways only reached titers that may be explained with passive product transport [26, 30, 32, 36, 37]. However, as demonstrated above, passive diffusion is certainly not suitable to achieve competitive product titers neither able to explain the titers reached in early works on production of muconic acid [16, 18, 78] in *E. coli* and the studies reporting routes via salicylate [29] and outgoing from lignin-derived species [21]. Also, the production of adipic acid with *Thermobifida fusca* [33] reached titers that indicate the presence of an active transporter in this species. Especially, for the titer of 2.1 $g \times L^{-1}$ in case of muconic acid production from dehydroshikimate in *S. cerevisiae* [25], an intracellular product concentration of ≈ 140 mM would be necessary for passive diffusion. This suggests the existence of some form of facilitated or active metabolite efflux in bacteria, but also in the eukaryote. Identification of the underlying export mechanism(s) could be a significant step towards more efficient muconic and adipic acid production, benefiting cases where intracellular product concentration restricts pathway thermodynamics. In addition, the analysis of intracellular metabolite concentrations may help answer the question of export limitation in thermodynamically favoured pathways.

1.2 Stoichiometric limitations of the different pathways

EMA showed a large spread of maximum theoretical carbon yield (Y_{max}) between the different routes (cf. figure 3 and supplementary files 2 & 3). The highest theoretical Y_{max} in both *E. coli* and *S. cerevisiae* were achieved by using the inositol pathway to produce glucaric acid. However, the few flux distributions that delivered up to 100% yield had no activity in central metabolism: only glucose phosphorylation, the production pathway and a part of the respiratory chain were active in these EFMs. This is almost equivalent to direct catalysis and may be physiologically impossible.

Routes of the 3-oxoadipate pathway came in second-best in both organisms in respect to their Y_{max}, reaching over 90%, directly yielding adipic acid. Here, the high yields are possible because the precursor, succinyl-CoA, can be derived from a reverse operating TCA cycle where no CO_2 is formed. However, CO_2 is formed in the course of the two routes, leading to carbon being lost (cf. section 2.2.3).

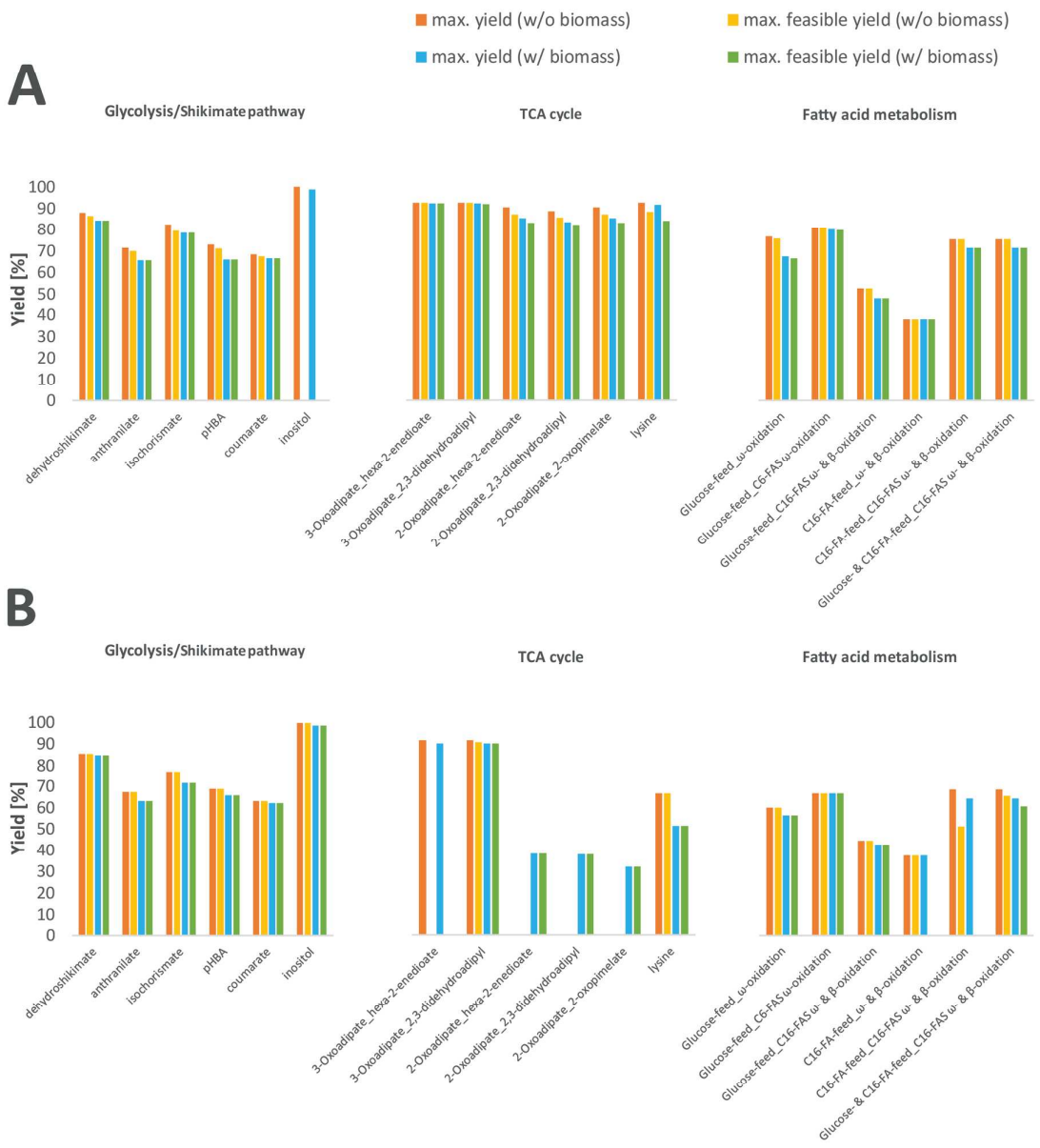


Figure 3: Maximum carbon yields for pathways to production of adipic acid (and its direct precursors) and impact of thermodynamics on these. Yields are given in % [C-mol/C-mol]. (A) *E. coli* and (B) *S. cerevisiae*.

In *E. coli*, the lysine degradation pathway achieved the same Y_{max} as the 3-oxoadipate pathways. However, the analogous route in *S. cerevisiae* has a Y_{max} of only 66%, due to (a) the partial mitochondrial localization of the pathway in yeast requiring additional energy for transport of metabolites and (b) 2-oxoglutarate initiating the pathway (similar to the 2-

oxoadipate pathway discussed below). Y_{max} of the 2-oxoadipate pathways were not much lower than 90% in *E. coli*; the yield only being slightly reduced due to formation of the precursor 2-oxoglutarate from TCA cycle involving production of CO_2. In *S. cerevisiae* however, the Y_{max} were very low and adipic acid production was stoichiometrically impossible without biomass formation. This means that product formation is coupled to growth and a direct competition exists, which resulted in a reduced product yield (cf. yield vs. biomass plots, supplementary files 4 & 5). The reason is that 2-oxoglutarate is a precursor for 2-oxoadipate pathways: unlike *E. coli*, *S. cerevisiae* does not have transhydrogenases, therefore the only major sink for cytosolic NADPH, which is produced in the cytosolic part of the TCA cycle during 2-oxoglutarate formation from isocitrate, is biomass.

The yields from routes based on the shikimate pathway (cf. section 2.2.1) ranged from 63% to 88%, with Y_{max} in general being slightly higher in *E. coli*. Y_{max} decreases the further downstream the route branches off from the shikimate pathway, mainly because additional bioconversions require more resources. For the routes via isochorismate a synergistic effect may be obtained when the two sub-routes via 2,3-didehydrobenzoate and salicylate are active, complementing their cofactor demands (one consumes, the other produces NADH), balancing it with central metabolism.

Routes based on FA metabolism (cf. section 2.2.6) showed a wide distribution of possible yields (figure 3), which greatly depended on the carbon-source and on the presence of anabolic and catabolic FA metabolism pathways. Here the highest Y_{max} in *E. coli* were obtained on a glucose-feed with direct synthesis of a C6-body (hexanoate) and subsequent ω-oxidation, while in *S. cerevisiae* the Y_{max} were highest from a C16-feed when combining FAS and β-oxidation. The overall lowest Y_{max} of 37.5% was obtained when adipic acid was produced from a C16-FA-feed via β-oxidation without FAS. This can be explained by loss of excess acetyl-CoA generated during β-oxidation of the C16-FA palmitate to the C6 body of

adipic acid (6 / 16 = 0.375): When FAS is implemented alongside β-oxidation, acetyl-CoA can be recycled into FA metabolism, thus benefiting product yield. The assumption that FA biosynthesis and oxidation could be metabolically active at the same time, however opposes natural transcriptional control [79] and substantial regulatory redesign of FA metabolism would be needed for simultaneous FAS and ω-, β-oxidation. It is noteworthy that a combined glucose- and C16-FA-feed could also not improve the Y_{max} significantly. In addition, it is worth mentioning that when assuming adipic acid to be the final product of β-oxidation, meaning it was only allowed to proceed to a C6-body instead of full breakdown of the FA into C2-bodys, a substantial minimum yield constraint of 37.5% was introduced (cf. yield vs. biomass plots, supplementary files 4 & 5).

Finally, there is value in taking into consideration that the shikimate pathway and FA metabolism based pathways are both limited to aerobic conditions due to the nature of the product pathways (oxygen is an essential substrate).

1.3 Pathway limitations and impact of compartment pH on Gibbs energies of reactions

For pathways with one or more $\Delta_rG'^0$ close to zero, a change in the compartment pHs due to environmental changes can impact a pathways thermodynamic feasibility. Further, the reaction Δ_rG at low and high ratios between products and substrates can provide an indication of a pathway's performance. Therefore, after establishing the differences in achievable carbon yields, the Δ_rG range for each reaction of the different networks were determined at different physiological conditions in a separate thermodynamic analysis (results are included in supplementary files 6 & 7). This thermodynamic evaluation provides an initial assessment of the pathways' general thermodynamic feasibility at different physiological conditions, in particular pH. In this context, the reader should be aware that the condition $\Delta_rG < 0$ is necessary but not sufficient for a reaction to occur *in vivo*. A reaction was considered thermodynamically feasible at a certain pH, if at least the Δ_rG_{min} was negative. If further also

the $\Delta_rG'^0$ was negative, the thermodynamic equilibrium lies on the product side and the reaction is thermodynamically favoured in the written direction. A metabolic route to adipic acid was considered favoured if these two criteria applied to all reactions on the route.

Determination of $\Delta_rG'^0$, Δ_rG_{min} and Δ_rG_{max} revealed a few routes with overall favourable thermodynamics, namely the dehydroshikimate, anthranilate and pHBA route in both *E. coli* and *S. cerevisiae*; all other pathways contained at least one reaction with a $\Delta_rG'^0 > 0$ at one or more of the tested physiological boundary pHs (outlined in Appendix A.1, cf. also "compartment" sheet in supplementary files 6 & 7). Further, both malate dehydrogenase reactions appeared to be irreversible in the forward direction at low pH conditions in *S. cerevisiae* (extracellular pH of 3.5). At low pH conditions in *E. coli* (extracellular pH of 5.9), the inositol pathway became feasible, while the dehydrogenase reaction in the lysine pathway remained infeasible. Adverse effects of low pH were seen in the 2- and 3-oxoadipate pathways, where the Δ_rG increased for most reactions. As expected, the impact was more pronounced in *S. cerevisiae*, as the yeast tolerates a much lower physiological pH.

In routes based on FA metabolism, the pH effect was mixed, as the thermodynamic equilibria of some reactions shifted towards favouring of product formation, while some turned in the other direction. Peroxisomal pH however had a profound impact on routes involving β-oxidation, which were only feasible at standard cytosolic pH (7.2) in combination with an alkaline peroxisome (pH 8.2) (cf. appendix A.1 for details about the different pHs). Although a high peroxisomal pH restricts the ATP/AMP antiporter, efflux of phosphate from the peroxisome is thermodynamically impossible at low peroxisomal pH (at least at 5.8 or lower, potentially also before). This supports studies arguing for an alkaline peroxisome [80] and the postulation of a proton gradient [81] that can potentially drive transporters.

1.4 Thermodynamic feasibility of elementary flux modes and impact on maximum carbon yields

NET-analysis was used to distinguish between thermodynamically feasible and infeasible EFMs. The $\Delta_r G'^0$ of all reactions were estimated with the exception of biomass reaction (see Appendix A.1). As the purpose of this study was to investigate production of adipic acid, only EFMs that included product formation were analysed. Considering thermodynamic feasibility severely reduced Y_{max} for some of the studied pathways (figure 3, for detailed information on the thermodynamic feasibility of the flux modes of each network refer to supplementary files 6 & 7 "infeasible-patterns" sheet). Infeasible flux distributions were identified in every network of each organism. Depending on the pathway, from 11% up to 100% of the product-forming EFMs in *E. coli* and from 5% up to 100% of the product-forming EFMs in *S. cerevisiae* were found thermodynamically infeasible (figure 4).

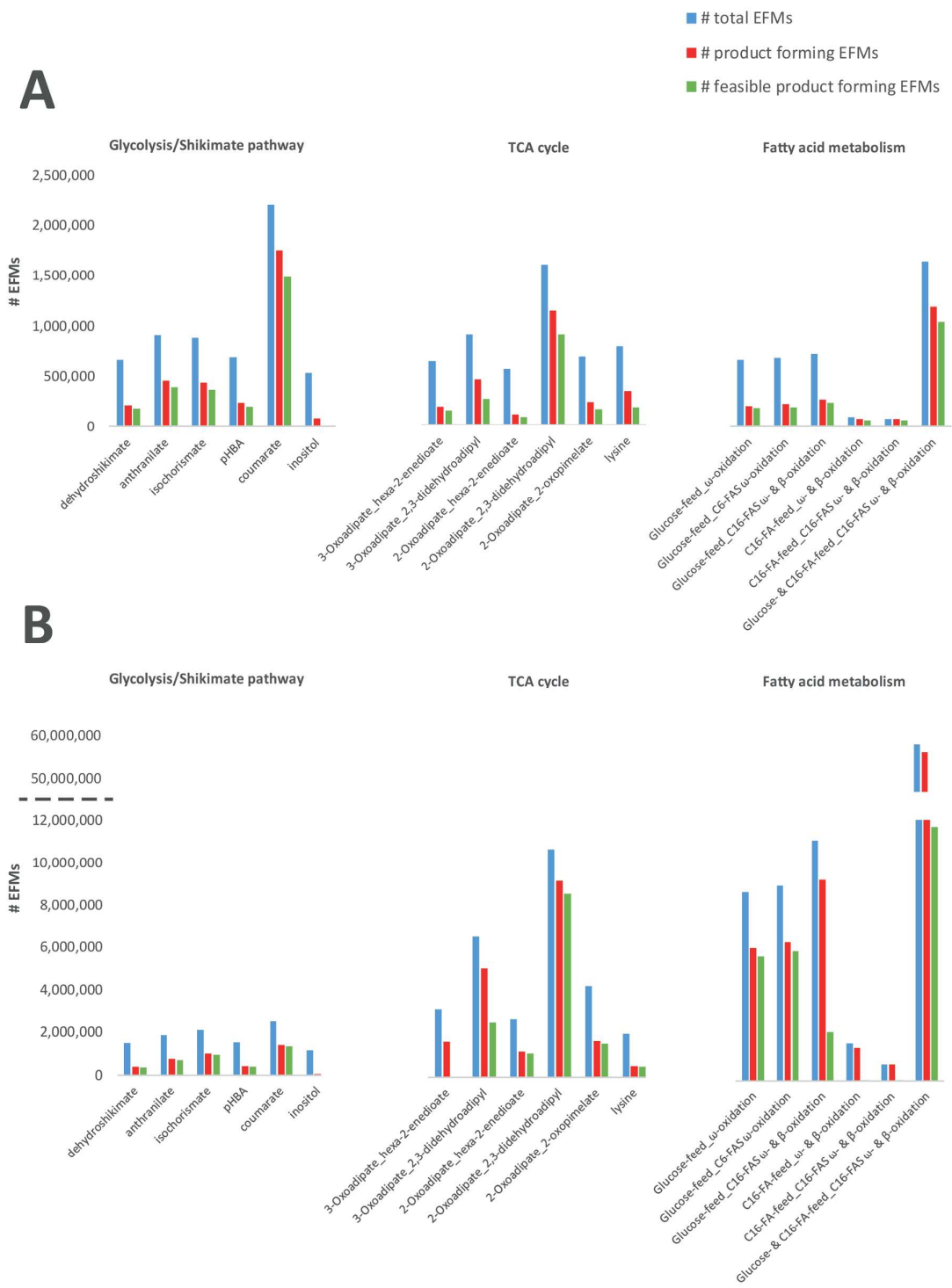


Figure 4: Numbers of EFMs for studied pathways to produce adipic acid, total EFMs (blue), product forming EFMs (red) and thermodynamically feasible EFMs (green), in (A) *E. coli* and (B) *S. cerevisiae*.

For the *S. cerevisiae* models, one *infeasible-pattern* rendered between 5 and 15% EFMs infeasible in every studied pathway. The *infeasible-pattern* is the combination of the mitochondrial malate dehydrogenase (R73) proceeding in forward direction (oxaloacetate production), and the malic enzyme (R74) proceeding in reverse direction (malate production). These two reactions are limited by the mitochondrial NAD/NADH ratio. For R73 to be feasible in forward direction or R74 to be feasible in reverse direction a minimum $\Delta_r G$ of 8.6 KJ/mol is required for the other reaction, due to the constraint on the MAL[m] × NAD[m]/NADH[m] ratio. Therefore, the combination of these two directionalities together is thermodynamically infeasible, although the malate dehydrogenase (*MDH1*) has been reported to be reversible and even favoured in forward direction (according to SGD [82]). These findings are in accordance with previous ones [63], where under restricted mitochondrial NAD and NADH concentrations the mitochondrial malate dehydrogenase was found to be infeasible in forward direction. However, the theoretical carbon yields were unaffected by this restriction (figure 3, cf. also supplementary file 5).

Infeasibility of the *S. cerevisiae* 3-oxoadipate pathway was caused by six additional patterns. The route via hexa-2-enedioate of the 3-oxoadipate pathway was completely infeasible due to contradicting restrictions of mitochondrial 3-oxoadipyl-CoA concentrations required by the succinyl-CoA:acetyl-CoA transferase (R115) and the 3-oxoadipyl-CoA transferase (R116). In order for R115 to proceed in forward direction, the maximum concentration of mitochondrial 3-oxoadipyl-CoA could be no higher than 0.12 mM. This did not comply with the subsequent reaction R116, which needed the metabolite at a minimum concentration of 0.33 mM in order to proceed in forward direction. The drop in maximum product carbon yield by 1% of the 2,3-didehydroadipyl-CoA route of the 3-oxoadipate pathway, was caused by a restriction of the mitochondrial ATP / (ADP × P) ratio, linked to the directions of R69 and R105, the mitochondrial succinyl-CoA synthetase and the mitochondrial ATP synthase. This pattern

caused the infeasibility of 41% of the modes of the route. There were further four complex thermodynamic *infeasible-patterns* affecting this route, all these patterns involved the triose-phosphate isomerase (R27), glyceraldehyde-3-phosphate dehydrogenase (R28), cytoplasmic malate dehydrogenase (R49), succinyl-CoA synthetase (R69), succinyl-CoA:acetyl-CoA transferase (R115), 3-hydroxyacyl-CoA dehydrogenase (R116) and the enoyl-CoA hydratase (R117). Unique to this pathway's routes are R116 and R117, which share the determining metabolite, 3-Hydroxyadipyl-CoA. As the Δ_rG^0 of R117 is close to zero (2.2 KJ/mol), the minimum concentration of 3-Hydroxyadipyl-CoA, which is required for R117 to proceed in forward direction, is 0.244 μM. However, the other reactions, including R116 and the ones individual to the four patterns, restrict the maximum concentration of 3-Hydroxyadipyl-CoA to 0.102 μM, resulting in infeasible EFMs (for details refer to supplementary file 7, "infeasible-patterns" sheet).

The fraction of infeasible EFMs in *S. cerevisiae* was greatest for FA based routes that involved the pathway of FA β-oxidation (76% or higher). Peroxisomal phosphate transport (R92) into the cytosol was partially responsible for this, thus only EFMs where phosphate could be balanced within the peroxisome were feasible. This was only possible in one of the three alternatives for release of adipate, namely release of adipate by the adipyl-CoA synthetase (EC 6.2.1.-) which allows peroxisomal phosphate to be recycled. However the Δ_rG_{min} of peroxisomal phosphate export appeared to be very close to 0 (< 1 KJ/mol, which is less than the error of the component contribution method [66]) and strongly dependent on pH (cf. section 1.3), therefore this infeasibility criterion may not be critical and the respective flux distributions may be feasible after all. Another pattern that affected the FA pathways that include β-oxidation required the reaction of the plasma membrane ATPase (R106) to proceed in the direction of ATP synthesis in around 25% of the product forming EFMs. This reaction is thermodynamically infeasible in that direction due to a Δ_rG_{min} of 21.9 KJ/mol. The last

pattern only affected networks where palmitate is a substrate and consisted of a reversed phosphoglycerate kinase (R29) and forward operating mitochondrial ADP/ATP translocase (R79) and mitochondrial ATP synthase (R105) reactions. In both networks that relied solely on palmitate-feed 98% of product forming EFMs were infeasible, while in the glucose and plamitate-feed co-feed network 9% of product forming EFMs were infeasible due to this pattern. The reasons for infeasibility were constraints on the cytosolic as well as mitochondrial ATP/ADP ratios: If R29 proceeds in reverse direction, the minimum ATP[c]/ADP[c] ratio is limited to 8.9, if in addition R79 proceeds in forward direction, the minimum ATP[m]/ADP[m] ratio is limited to 17.3. This conflicts with the maximum ATP[m]/ADP[m] ratio of 1.4 if R105 proceeds in forward direction, and maximum ATP[c]/ADP[c] ratio of 0.7 if also R79 proceeds in forward direction.

For *E. coli,* the different physiological conditions and range of metabolite concentrations led to other patterns constraining network thermodynamics: eight *infeasible-patterns* were found, but only four of them generated restrictions where the Δ_rG_{min} was far enough from zero to be considered significant. For details on all *E. coli infeasible-patterns* refer to supplementary file 7, sheet "infeasible_patterns". The restrictions reduced the number of EFMs feasible under the current conditions and therefore the feasible Y_{max} (cf. product vs. biomass yield plots, supplementary file 4) of twelve out of the eighteen studied routes (glycolysis based pathways, routes via the 2-oxoadipate pathway, the lysine pathway and the direct ω-oxidation as well as C6-FAS ω-oxidation pathways). Moreover, in six out of the twelve affected routes the maximum yield that included formation of biomass was also reduced (figure 3).

Four pathways were affected by thermodynamic restrictions: all product-forming EFMs of the inositol pathway were infeasible in *E. coli* due to the thermodynamic equilibrium of the uronolactonase reaction (R87), which was on the substrate side. This is surprising as an *in vivo* route for production of glucaric acid in *E. coli* has been reported previously, reaching

significant titers in the g/L range [35]. One explanation is that the uronolactonase reaction is energetically coupled with another reaction, which is especially likely as all other reactions in the pathway are thermodynamically greatly favoured. The first description of the pathway, where the two final steps, the uronolactonase and the aldehyde dehydrogenase reaction, are catalysed by a single gene product [83] supports this hypothesis. Additionally, according to KEGG [84] an alternative branch of the inositol pathway may exist, which would by-pass the problematic reaction via gulonate.

Further, 32% of the product forming EFMs of the *E. coli* 2-oxopimelate route, part of the 2-oxoadipate pathway, were infeasible (figure 4). A pattern, unique to this route, was the cause for half of the total infeasible EFMs of this route. The three contradicting constraints of the pattern were: the malate dehydrogenase (R56) proceeding in forward direction, the NADP dependent malic enzyme (R61) proceeding in reverse direction and R85 proceeding in forward direction (for more details on the *infeasible-pattern* refer to supplementary file 6). Similarly, the 2,3-didehydroadipyl-CoA route of the 3-oxoadipate pathway had an *infeasible-pattern* responsible for half of the infeasible product forming EFMs of this route. The pattern, restricted by the ratio between succinyl-CoA and coenzyme A, consists of three reactions: a reverse operating succinyl-CoA synthetase (R53) and forward operating ATP synthase (R79) and succinyl-CoA:acetyl-CoA transferase (R84). Despite 43% of the EFMs of this route being infeasible, the Y_{max} was not affected (figure 3). Product forming EFMs of the pathway to adipic acid via lysine were also partially (47%) infeasible in *E. coli*, reducing the feasible Y_{max} to 88% and the maximum yield with biomass formation to 84%. The main reason was the non-enzymatic reaction in the dehydrogenase branch of the lysine biosynthesis (R94 in the respective network, KEGG [84] reaction number R04336), which was infeasible under the studied conditions (affecting 35% of the EFMs of this pathway) although many organisms, like e.g. *C. glutamicum*, include this biosynthetic pathway. This might explain the parallel

presence of the succinylase branch and why it is exclusive in many organisms, like e.g. *E. coli*, although the achievable carbon yield is slightly lower.

Thermodynamic feasibility-assessment of novel pathways leading to the biosynthesis of traditionally chemically produced compounds has previously been done using thermodynamic metabolic flux analysis (TMFA) [85]. TMFA directly estimates metabolic fluxes taking into account consistency of flux directionality with thermodynamic constraints [86]. Because TMFA relies on metabolic flux analysis, the result is narrowed down to the solution of maximal product formation. Thus, an incomplete picture of the organism's metabolism is drawn. As a way to describe all metabolic steady states, EMA determines the full metabolic solution space of a biochemical network [87]. The present approach relies on EMA, therefore providing the whole picture of the thermodynamically feasible metabolism of the studied organism.

As demonstrated, *NExT-EMA* provides additional insight into metabolism as it allows identification of the source of infeasibility of certain flux distributions, including the reactions involved, and a deep understanding of the cause and consequences of these *infeasible-patterns*. These can be understood in a similar way as with the previously published *tEFMA* for calculation of thermodynamically feasible EFMs, which is also able to identify infeasible-patterns and can deliver insight into causes of infeasibilities [64, 88]. However, with *tEFMA*, a change in for instance concentration boundaries of metabolites requires a complete new run while *NExT-EMA* requires only re-analysis of the thermodynamic part, which may be faster. Also, with *tEFMA*, it is unknown, which EFMs are lost, as only minimal infeasible patterns can be tracked, while with *NExT-EMA* both, infeasible EFMs and infeasible patterns are determined. By knowing the combinations of reactions that cause infeasibility, a strain engineering strategy can be developed that avoids these reactions. Putting it the other way around, this study will not only help to choose the

best pathway but also aid in the development of tailored strain design strategies. Further, the tool is not limited to production of adipic acid and the same principles and analyses can be applied to any microbial production system at any stage of the project. Together with the most recent implementation for enumeration of EFMs, *FluxModeCalculator* [89], which significantly reduces computational times and costs, *NExT-EMA* enables the analysis of very large metabolic networks on desktop-computers.

1.5 Pathway robustness and consequences for commercial application

In order to allow an assessment of the pathways regarding industrial applicability and to give a recommendation regarding the highest potential for economic viability, a study under tightened conditions was conducted (industrial viability analysis). The particular conditions were chosen in respect to minimization of substrate and purification costs (minimum carbon yield of 75% and product titer of 100 mM, respectively), and productive use of capital infrastructure (minimum rate of 0.01 mol/(g_{DW}×h)). Based on published values [90, 91], these industry requirements were slightly adjusted (cf. appendix A.2), to account for limitations implied by the particular target product(s) in the present case study. The minimum rate and enzyme kinetics assumptions were used to narrow the metabolite concentrations ranges (cf. appendix A.2 and supplementary file 8 for details on assumptions and calculations). However, under these conditions no feasible EFMs were found in either organism. When incrementally relaxing the constraints on minimum carbon yield and product titer, two alternative scenarios were found: a high minimum product carbon yield (75% or higher) is only achievable at a low minimum titer (5 mM), and vice versa – a high minimum titer (100 mM) is only possible in pathways that deliver lower yields (i.e. when the minimum product yield constraint has been dropped to 50%). Multiple additional *infeasible-patterns* existed, reducing the number of thermodynamically feasible EFMs, which impacted the *S. cerevisiae* models more severely than the *E. coli* ones (cf. supplementary file 9). In detail, just the

dehydroshikimate and isochorismate routes in *E. coli* proved to be able to deliver external product at a concentration of 5 mM or higher and a carbon yield of 75% or higher (figure 5), while solely in the dehydroshikimate route EFMs existed, which also allowed formation of biomass. Therefore, the *E. col*i dehydroshikimate route of the shikimate pathway has the greatest potential for bio-based production of adipic acid. This correlates with the fact that to-date the by far highest reported yields (22% & 44%) were achieved via the dehydroshikimate route in *E. coli* [16, 92]. A high minimum titer of 100 mM was possible in the 2,3-didehydroadipyl-CoA route of the 3-oxoadipate pathway, all routes of the 2-oxoadipate pathway (including the 2-oxopimelate and lysine pathway), as well as the direct ω-oxidation and the FA pathways that included C16-FAS in *E. coli*, but only at product carbon yields higher than 50% but lower than 75%. In *S. cerevisiae* two pathways remained feasible at the 50% minimum product carbon yield and a 100 mM minimum product titer scenario: these were the coumarate route and the pathway of direct ω-oxidation, while in the former just 22 of 14,473 EFMs remained feasible (which may not represent enough metabolic freedom for the organism to be viable *in vivo*, as these EFMs feature almost the same flux distribution and above all do not produce biomass). Further, the range of feasible metabolite concentrations of the remaining feasible EFMs appeared to be very narrow, indicating that a careful balancing of the pathways will be pivotal to achieve these flux distributions. All results are compiled in supplementary file 9, where also the distribution of feasible EFMs by means of product vs. biomass yield plots can be found; the feasible metabolite concentrations for the respective scenarios can be found in supplementary file 10.

The outcome of this analysis once again demonstrates the potential of *NExT-EMA*: the solution space can be narrowed down by more specifically defining the desired target compound production parameters and the range of metabolite concentrations. In that way, it may also become possible to elicit the plausible *in vivo* flux distribution of a certain

metabolism without actually having to measure the metabolic fluxes: when investigating a particular organism *in vivo*, experimentally acquired metabolite data can be used to generate constraints that may limit the solution space of a corresponding model. The result would be a narrow distribution of feasible EFMs, which reflects the state(s) present *in vivo*.

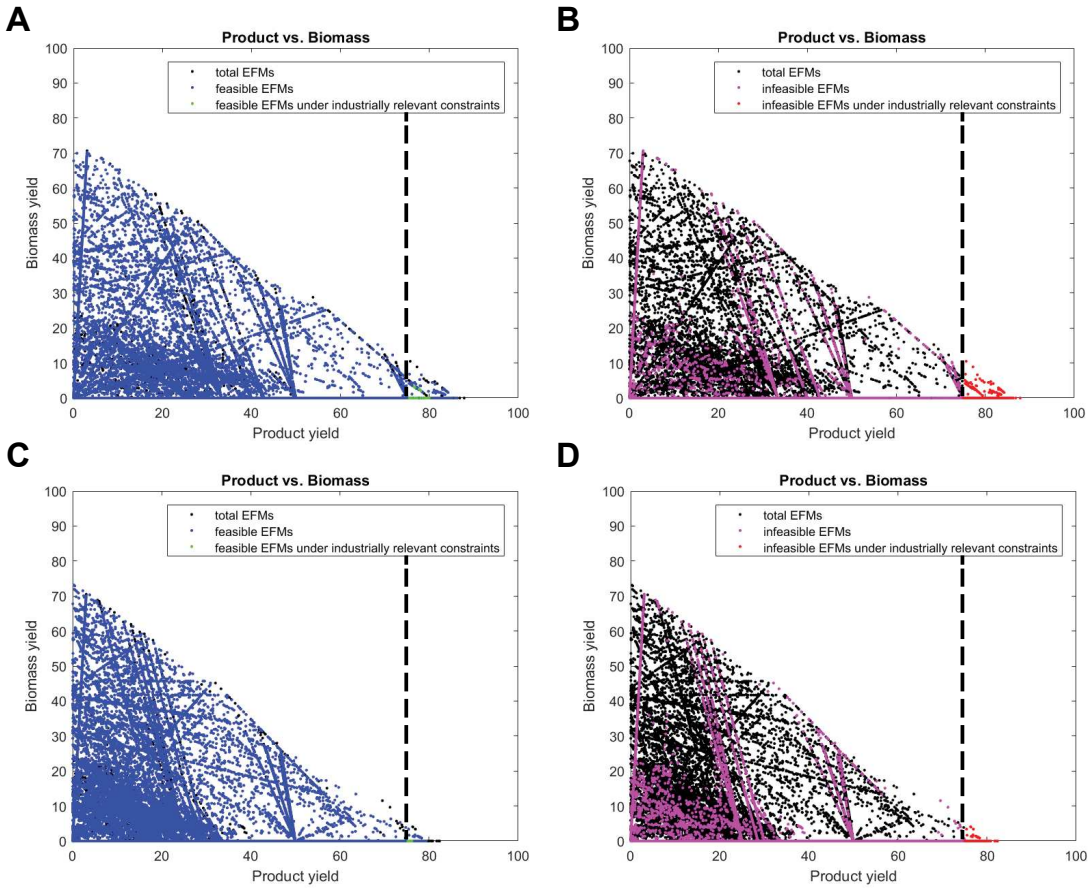


Figure 5: Product vs. biomass carbon yield plots of feasible and infeasible modes from NExT-EMA analysis of certain shikimate pathway based routes in *E. coli*. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %. Color code: feasible EFMs under physiological conditions (metblue), feasible EFMs under industrially relevant conditions (green), infeasible EFMs under physiological conditions (pink), infeasible EFMs under industrially relevant conditions (red). (A) Feasible EFMs of the dehydroshikimate route, (B) Infeasible EFMs of the dehydroshikimate route, (C) Feasible EFMs of the isochorismate route, (D) Infeasible EFMs of isochorismate route.

It is important to keep in mind that the full solution space is made up of the elementary flux modes and their linear combinations. When considering the thermodynamically feasible EFMs, it is, however, not a given that all linear combinations of these are also always feasible [63]. This is especially important for the understanding of figure 5, where the feasible EFMs may or may not span an area greater than the single feasible EFMs, in which metabolism may operate.

2. Methods

2.1 Assessment of cellular transport processes

Different scenarios for the export of dicarboxylic acids from the cytosol to the medium were investigated in detail. In particular, passive transport via diffusion, proton mediated symport, transport of fully dissociated acid and ATP driven export by ABC-type transporters were compared. Passive transport relies on the free diffusion of small, uncharged molecules over the cell membrane. The minimal total internal acid concentration, c_{int}, required to transport against a given total external concentration, c_{ext}, is determined as the concentration where internal and external un-dissociated forms are in equilibrium. In general:

$$\frac{c_{int}}{c_{ext}} = \frac{[H_2A]_{int}}{[H_2A]_{ext}} \times \frac{1 + (10^{pH_{int} - pK_{a1}}) \times (1 + 10^{pH_{int} - pK_{a2}})}{1 + (10^{pH_{ext} - pK_{a1}}) \times (1 + 10^{pH_{ext} - pK_{a2}})} \quad (1)$$

Considering that the internal and external un-dissociated forms are in equilibrium ($[H_2A]_{int}=[H_2A]_{ext}$), the ratio between internal and external total acid concentration can be expressed as:

$$\frac{c_{int}}{c_{ext}} = \frac{1 + (10^{pH_{int} - pK_{a1}}) \times (1 + 10^{pH_{int} - pK_{a2}})}{1 + (10^{pH_{ext} - pK_{a1}}) \times (1 + 10^{pH_{ext} - pK_{a2}})} \quad (2)$$

For a proton anionic symporter of the fully dissociated form of a dicarboxylic acid Eq. (3) can be formulated (derived from [68]), which includes a factor that accounts for the efflux of protons:

$$\frac{c_{int}}{c_{ext}} = 10^{n \times \Delta pH} \times \frac{1 + (10^{pK_{a2} - pH_{int}}) \times (1 + 10^{pK_{a1} - pH_{int}})}{1 + (10^{pK_{a2} - pH_{ext}}) \times (1 + 10^{pK_{a1} - pH_{ext}})} \quad (3)$$

where n is the number of transported protons (in case of a dicarboxylic acid n equals 2) and ΔpH the difference in pH across the membrane (pH_{int} - pH_{ext}).

The third studied case is the transport of a fully dissociated acid, which is not charge-balanced and requires a concomitant transport of protons to maintain the pH homeostasis. The proton transport is mediated by an ATPase. Therefore, two simultaneous reactions take place:

$A^{2-}_{in} = A^{2-}_{out}$ and $2\ H^{+}_{in} + 2 \times f \times ATP = 2\ H^{+}_{out} + 2 \times f \times ADP + 2 \times f \times P_i$

where f is the ATP cost for transportation of one proton (number of ATP molecules needed to export one proton). The most optimistic case of energy expense is to assume the reactions fully coupled, resulting in the following ratio of internal to external concentration of the acid:

$$\frac{c_{int}}{c_{ext}} = 10^{n(\Delta pH + f \times \frac{\Delta G_{ATP}}{ZxF})} \times \frac{1 + (10^{pK_{a2} - pH_{int}}) \times (1 + 10^{pK_{a1} - pH_{int}})}{1 + (10^{pK_{a2} - pH_{ext}}) \times (1 + 10^{pK_{a1} - pH_{ext}})} \quad (4)^1$$

where ΔG_{ATP} is the energy generated by ATP hydrolysis (ATP + H_2O → ADP + P_i), calculated using *NExT* [61] and adjusted for the properties of the respective compartments.

The per-ATP-numbers of protons exported across the cytosolic membranes of the different organisms is open for debate [93]. In accordance with [94, 95], we assumed 3.3 and 1 for *E. coli* and *S. cerevisiae*, respectively. Thus, 0.6 and 2 ATP molecules are used per dicarboxylic acid transported for *E. coli* and *S. cerevisiae*, respectively.

In case of active transport of the un-dissociated acid driven by ATP hydrolysis, the minimal intracellular concentration required to achieve diacid transport is adjusted by a factor accounting for the additional driving force of ATP hydrolysis [68], which is similar to the

[1] Z = ln(10)×T×R/F, while R (gas constant) = 8.31446 J/(mol×K) and F (faraday constant) = 96,485.3329 A×s/mol. T depends on the (optimum) cultivation temperature (in Kelvin) of the respective microorganism.

previous case but with a fixed ratio of 1 ATP consumed per dicarboxylic acid transported, resulting in Eq. (5):

$$\frac{c_{int}}{c_{ext}} = \left(10^{\left(\frac{\Delta G_{ATP}}{Z \times F}\right)}\right) \times \frac{1 + (10^{pH_{int} - pK_{a1}}) \times (1 + 10^{pH_{int} - pK_{a2}})}{1 + (10^{pH_{ext} - pK_{a1}}) \times (1 + 10^{pH_{ext} - pK_{a2}})} \quad (5)$$

Detailed step-by-step calculations of the different product export scenarios can be found in supplementary file 1.

2.2 Biochemical pathways to adipic acid

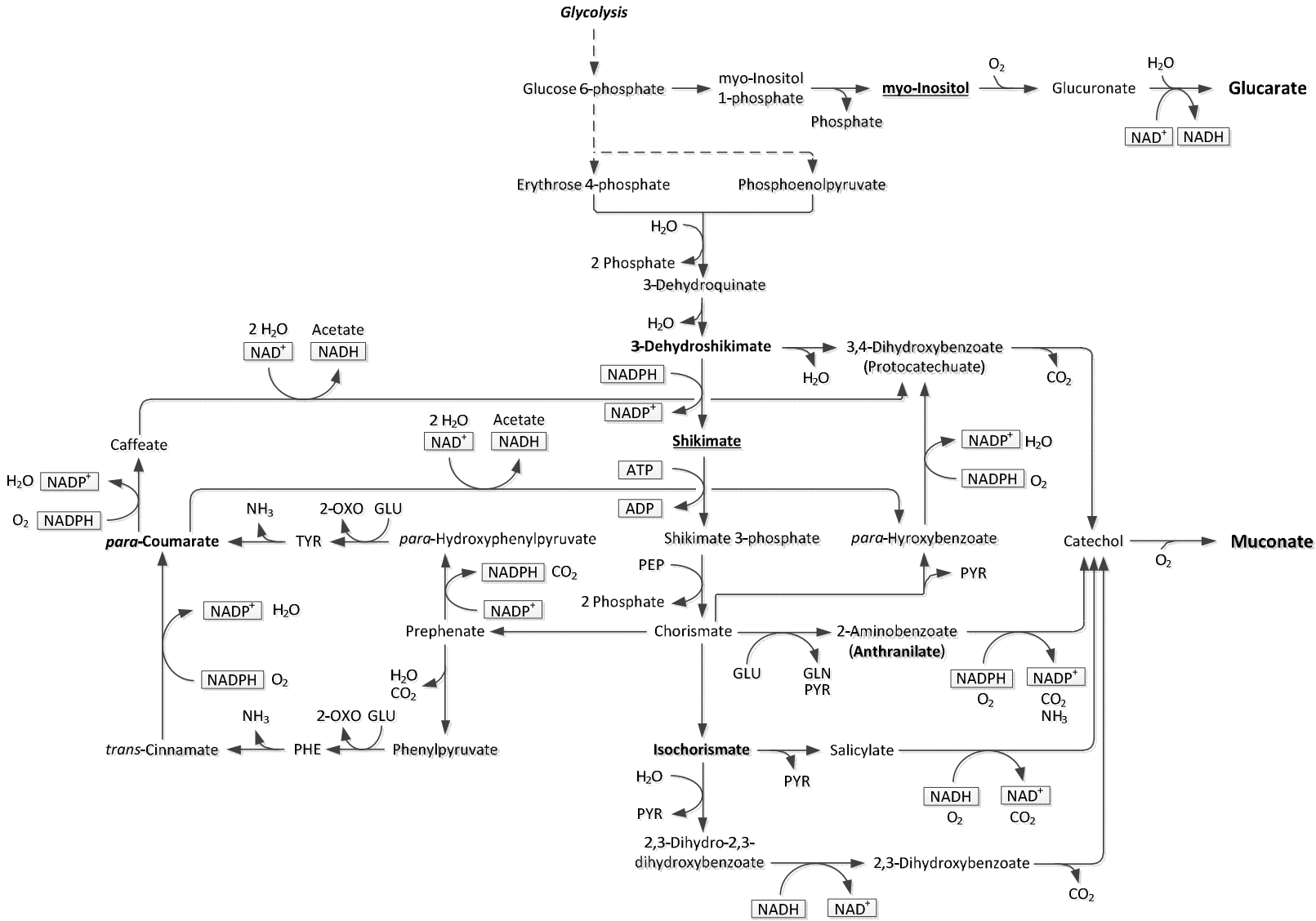
Six distinct pathways featuring unique intermediates and enzymatic conversions were compiled from patents [18, 19, 39-43] and scientific publications [26-34, 36, 37, 78]. In total 16 different routes for carbon flow were considered. Each route was implemented in two different organism models: *E. coli* and *S. cerevisiae* (cf. supplementary files 2 & 3 and 6 & 7).

2.2.1 Muconic acid production via the shikimate pathway

It is difficult to define separate routes for the production of *cis,cis*-muconic acid from the shikimate pathway as many options exist regarding host organism, biochemical route, heterologous enzymes, carbon-source and fermentation method [96]. Here, five different routes were distinguished; three of these contain alternative inter-conversions (figure 6). The routes were named according to key intermediates / branch off from shikimate pathway: 3-dehydroshikimate, anthranilate, isochorismate, pHBA (*para*-hydroxybenzoate) and *para*-coumarate.

2.2.2 Glucaric acid production via inositol

D-glucaric acid can be derived from inositol metabolism. From the central metabolite glucose-6-phosphate, branch point of glycolysis and pentose phosphate pathway, myo-inositol is formed in two reactions. In two further reactions D-glucaric acid is obtained (figure 6).


Glycolysis
Glucose 6-phosphate
myo-Inositol 1-phosphate
Phosphate
myo-Inositol
O2
Glucuronate
H2O
Glucarate
NAD+
NADH
Erythrose 4-phosphate
Phosphoenolpyruvate
H2O
2 Phosphate
3-Dehydroquinate
H2O
3-Dehydroshikimate
3,4-Dihydroxybenzoate (Protocatechuate)
H2O
CO2
NADPH
NADP+
2 H2O
Acetate
NAD+
NADH
Caffeate
Shikimate
ATP
ADP
NADP+ H2O
NADPH O2
H2O NADP+
O2 NADPH
NH3
2-OXO GLU
para-Coumarate
TYR
para-Hydroxyphenylpyruvate
Shikimate 3-phosphate
para-Hyroxybenzoate
Catechol
Muconate
O2
PEP
2 Phosphate
PYR
NADPH CO2
NADP+
NADP+ H2O
Prephenate
Chorismate
2-Aminobenzoate (Anthranilate)
NADPH O2
H2O CO2
GLU
GLN PYR
NADPH O2
NADP+ CO2 NH3
NH3
2-OXO GLU
trans-Cinnamate
PHE
Phenylpyruvate
Isochorismate
Salicylate
PYR
H2O
PYR
NADH O2
NAD+ CO2
2,3-Dihydro-2,3-dihydroxybenzoate
2,3-Dihydroxybenzoate
CO2
NADH
NAD+

621

Figure 6: Glycolysis based pathways for biological production of the adipic acid precursors muconic and glucaric acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

2.2.3 Production via the 3-oxoadipate pathway

The 3-oxoadipate pathway, previously reported as "reverse adipate degradation" [32, 36, 41, 42], leads over two different routes to the intermediates hexa-2-enedioate and 2,3-didehydroadipyl-CoA (figure 7). From there, the former is directly converted to adipic acid, the latter is first reduced to adipyl-CoA. For the conversion of adipyl-CoA to adipic acid, two options were explored: one step release of adipic acid in a hydrolase reaction via an adipyl-CoA thioesterase[36] or two steps utilizing promiscuous phosphate butyryltransferase and butyryl kinase enzymes[32]. As the pathway intermediate, succinyl-CoA, is not a cytosolic metabolite in *S. cerevisiae* the pathway was assumed to be mitochondrial in the yeast model.

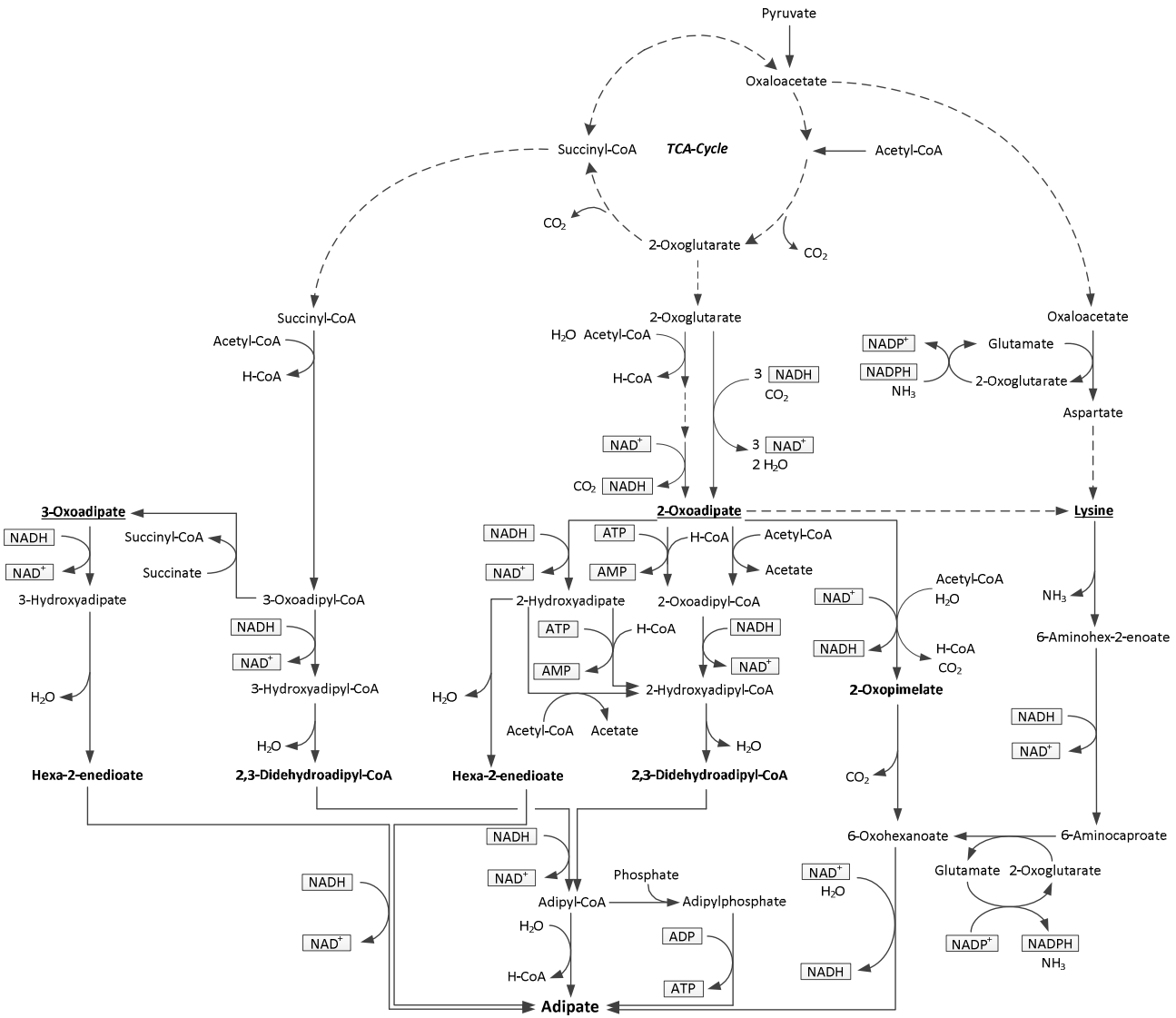
2.2.4 Production via the 2-oxoadipate pathway

As for the 3-oxoadipate pathway, multiple routes have been described for the production of adipic acid via 2-oxoadipate [41, 42], partly involving the same metabolites as the 3-oxoadipate pathway. All routes were compiled into one interlinked pathway (figure 7) and nine different possible routes for carbon flow were identified for analysis, in accordance with the respective patents. For the conversion of adipyl-CoA to adipic acid the same reactions as for the 3-oxoadipate pathways were assumed. It should be mentioned that while each route is essentially unique in its steps of biochemical conversions, they may share the same overall stoichiometry. Therefore, carbon yields can be identical, while thermodynamics may differ.

2.2.5 Production via lysine degradation

Formation of adipic acid from lysine (figure 7) is closely related to 2-oxoadipate metabolism. In most bacteria, lysine is synthesized from aspartate and broken down via 6-oxohexanoate, which enters the 2-oxoadipate pathway. Further, in most prokaryotes, including *E. coli*, production of lysine from aspartate proceeds via a succinylase, while others, like *C. glutamicum*, in addition to the succinylase also contain a dehydrogenase [97]. Both options were included in the *E. coli* model to broaden the universality of the study. Yeasts synthesize

lysine naturally via 2-oxoadipate partially in the mitochondria, rather than through the aspartate biosynthesis. Therefore, only the last steps of this pathway coincide in the prokaryotic and eukaryotic models.


Pyruvate
Oxaloacetate
Succinyl-CoA
TCA-Cycle
Acetyl-CoA
CO2
2-Oxoglutarate
CO2
Succinyl-CoA
Acetyl-CoA
H-CoA
2-Oxoglutarate
H2O Acetyl-CoA
H-CoA
3 NADH
CO2
NAD+
3 NAD+
2 H2O
CO2 NADH
Oxaloacetate
NADP+
Glutamate
NADPH
NH3
2-Oxoglutarate
Aspartate
3-Oxoadipate
NADH
NAD+
Succinyl-CoA
Succinate
3-Hydroxyadipate
3-Oxoadipyl-CoA
NADH
NAD+
3-Hydroxyadipyl-CoA
H2O
H2O
Hexa-2-enedioate
2,3-Didehydroadipyl-CoA
2-Oxoadipate
Lysine
NADH
NAD+
ATP
AMP
H-CoA
Acetyl-CoA
Acetate
2-Hydroxyadipate
2-Oxoadipyl-CoA
ATP
AMP
H-CoA
NADH
NAD+
2-Hydroxyadipyl-CoA
H2O
Acetyl-CoA
Acetate
H2O
Hexa-2-enedioate
2,3-Didehydroadipyl-CoA
NAD+
Acetyl-CoA
H2O
NADH
H-CoA
CO2
2-Oxopimelate
CO2
NH3
6-Aminohex-2-enoate
NADH
NAD+
6-Oxohexanoate
6-Aminocaproate
Glutamate
2-Oxoglutarate
NADP+
NADPH
NH3
NADH
NAD+
Phosphate
Adipyl-CoA
Adipylphosphate
H2O
H-CoA
ADP
ATP
NAD+
H2O
NADH
Adipate

654

Figure 7: TCA cycle based pathways for biological production of adipic acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

2.2.6 Production via fatty acid metabolism

A unique variant to obtain adipic acid from engineered microorganisms is through fatty acid (FA) metabolism. Inferred from the molecular structure this seems natural as adipic acid consists of a saturated C_6-backbone with terminal α,ω-carboxy groups, making it basically a short chain dicarboxylic fatty acid. Rather than distinguishing between different routes this variant of adipic acid production is determined by the type of carbon-source used and the point of entry into the pathway cascade: according to a patent [40], both fatty acids and glucose can be substrates. If glucose is the sole substrate, synthesis proceeds either directly from central metabolism through ω-oxidation to adipate, or via fatty acid synthesis (FAS) and subsequent breakdown of the FAs through β-oxidation and carboxylation via ω-oxidation. If FAs are the carbon-source, these can directly enter β-oxidation and proceed through ω-oxidation to adipic acid. For the pathways described above (figure 8) a mixture of FAs is utilized either solely or in combination with glucose for production of adipic acid in *Candida tropicalis* [40] (production of dicarboxylic acids has recently been modelled in *C. tropicalis* [98]). We transferred the pathway(s) to a standard model of the budding yeast *S. cerevisiae* for more universal significance and comparability with the other pathways. Moreover, the composition of the FA mixture is not explicitly defined in the patent and might vary between batches. Therefore, the most common saturated FA palmitate (C_{16}) was assumed as a representative of the FA mixture. For the conversion of adipyl-CoA to adipate three possibilities were examined according to described enzyme classes [41, 42]: an adipyl-CoA hydrolase (EC 3.1.2.-), an adipyl-CoA synthase (EC 6.2.1.-) / phosphotransadipylase/adipate kinase (EC 2.3.1.-/2.7.2.-) and an adipyl-CoA:acetyl-CoA transferase (EC 2.3.1.-).

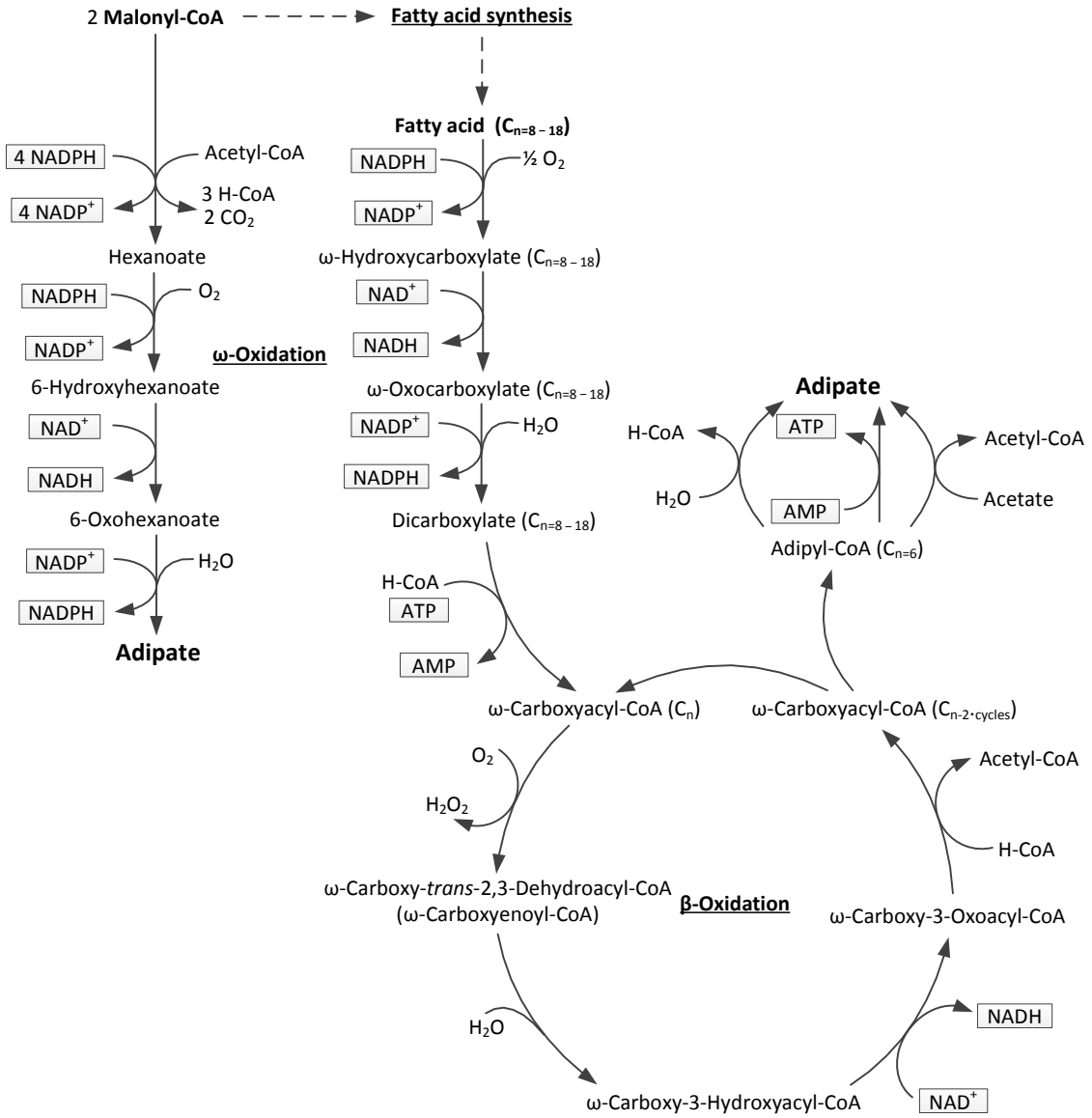


Figure 8: Fatty acid metabolism based pathways for biological production of adipic acid. Different pathways are written in bold and are underlined. Rather than distinguishing between different routes to adipic acid this variant for adipic acid production is determined by the type of carbon-source used (not apparent from the figure) and the points of entry into the pathway cascade, which are indicated in bold. Dashed lines indicate multiple steps.

2.3 Stoichiometric networks of the central carbon metabolism

Metabolic networks of *E. coli* and *S. cerevisiae* were based on previously published ones [99]. In order to enable FA metabolism based production pathways, the yeast network was

expanded with a peroxisomal compartment [100] and additional transport reactions [81], while palmitate uptake as an alternative substrate was added to both networks.

Biophysical properties (pH, ionic strength, redox potential, relative volume) of the different compartments (extracellular, cytosol, mitochondrial matrix, intermembrane space and peroxisome) were largely adopted from previous analyses [61]. For further details on principles, (physical) data and assumptions for construction of models, the reader is referred to appendix A.1.

2.4 Determination of feasibility limits in dependence of pH

The metabolic networks and pathways were subject to a preliminary and separate NET-analysis at different pH conditions in order to evaluate the thermodynamic equilibria of the single reactions of the different pathways (results are included in supplementary files 6 & 7). The pH can shift thermodynamic equilibrium either direction, depending on which side of an equation the protons take part in, and for transport reactions in which direction a proton gradient exists. In order to investigate this, the Gibbs free energies at lowest feasible physiological pH values were studied (*S. cerevisiae* extracellular: 3.5, cytosol: 4.4 [101, 102]; *E. coli* extracellular: 5.9 cytosol: 6.1 [103]). The pH of other compartments was scaled relatively to the cytosol, e.g. where no mitochondrial pH was determined, a constant ΔpH of + 0.3 in respect to the cytosol was assumed. To capture the boundaries of thermodynamic feasibility, the widest constraints were applied, and all reactions were considered reversible. Based on this, also Gibbs free energies at boundary reactant concentrations (Δ_rG_{min} and Δ_rG_{max}) were determined for each reaction in addition to the transformed standard Gibbs free energy ($\Delta_rG'^0$). The Δ_rG_{min} was determined using highest substrate and lowest product concentrations and the Δ_rG_{max} using lowest substrate and highest product concentrations, analogous to what has been described previously [61]. However, if a single reaction had a

positive $\Delta_r G_{min}$ the whole route was considered infeasible at the studied pH. All routes that were neither favoured nor infeasible were classified thermodynamically restricted.

2.5 Network-embedded thermodynamic analysis of elementary flux modes

EFMs can be computed with *EFMTool* [104] or other implementations for calculation of flux modes, e.g. the recent implementation *FluxModeCalculator* [89], which enables EMA of large scale metabolic networks. We trialled both, the use of *EFMTool* to calculate the flux distributions, or using *EFMTool* to only parse the networks into a stoichiometric matrix, which then served as input for *FluxModeCalculator* to calculate binary EFMs. The EFMs were then analysed with NET-analysis, using a modified version of *NExT* [67] (tool available on SourceForge, https://sourceforge.net/p/next-ema/); the basic workflow is outlined in **Error! Reference source not found.**. NET-analysis was performed at physiological conditions (cf. Appendices A.1 and A.2) and was limited to modes that involved formation of the product of interest (modes with no flux to and thus zero yield of adipic, muconic or glucaric acid are irrelevant in the present study). Computations were performed with MATLAB (The MathWorks, Natick, USA) using the built in fmincon solver and the ILOG CPLEX optimizer (IBM, New York, USA) on a desktop computer (Intel Core i7-3770, 32 GB RAM, 256GB SSD).[2] The maximum carbon yields of the different EFMs were determined with MATLAB by drawing carbon balances around the transport reactions in and out of the network (scripts included with program files).

[2] Enumeration of EFMs proceeded at a minimum speed of 3,000 modes/sec when using the *EFMTool* and at an average speed of 50,000 modes/sec when using the *FluxModeCalculator* (van Klinken & Willems van Dijk, 2015). *NExT* reached a minimum speed of 120 modes/sec in case of the most complex *S. cerevisiae* network, when relying on the fmincon solver. A different solver, as for example LINDO™, may accelerate the process significantly, especially in combination with parallel computing. In case the number of EFMs of a network becomes too large to fit into memory for at once analysis, sizeable batches can be defined to allow incremental analysis.

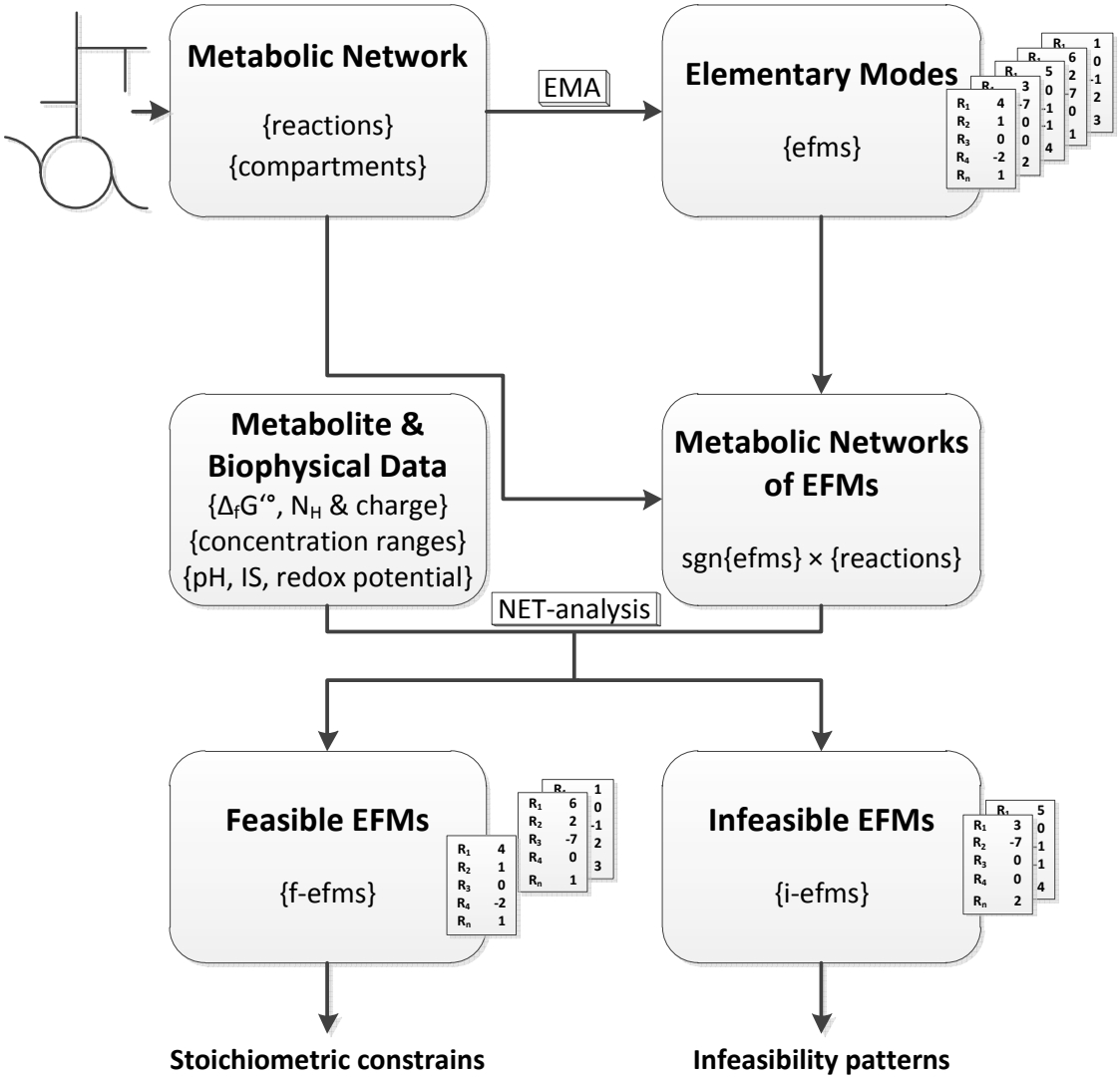


Figure 9: Flowchart for NET-analysis of elementary flux modes. EMA = elementary mode analysis; NET-analysis = network-embedded thermodynamic analysis; IS = ionic strength; sgn = sign function: If EFMs are not already binary, reactions were constrained according to the direction (≙ sign) of the flux determined, for zero flux the reaction was considered unconstrained.

The results of the thermodynamic-based network analysis of the EFMs of the different networks were further analysed. In order to find the cause of thermodynamic infeasibility of EFMs, feasible and infeasible EFMs were compared (cf. appendix A.3 for details). Briefly, a two-step algorithm was used: (1) identification of reaction constraint/s that if removed, would render the studied model thermodynamically feasible (2) verification that the directionality in

the particular reaction/s was/were sufficient to render the model infeasible. These two steps were repeated, introducing further reactions, until the smallest set of reactions, which was required to render the model infeasible, was found. This set of reaction directionalities was termed an *infeasible-pattern*. Finally, the infeasible EFMs were crosschecked in order to verify that removing the identified pattern constraints would cause them to become feasible. If infeasible EFMs remained after the removal of the pattern constraints, the process was repeated with the remaining infeasible EFMs in order to find further *infeasible-patterns*.

The analysis of the feasible and infeasible EFMs, the identification of *infeasible-patterns* and the final check of the infeasible EFMs were scripted in MATLAB (scripts included with program files, https://sourceforge.net/p/next-ema/).

3. Conclusions

Known metabolism currently offers 16 different solutions to the problem of biological adipic acid production. Ten of these routes lead directly to adipic acid, while six routes would yield the precursors *cis,cis*-muconic or glucaric acid. Here, we present the first comprehensive evaluation of differences in theoretical yields and thermodynamic feasibility of these pathways.

Analysing different product export scenarios, it was evident that high efflux is only possible at expense of a biological energy carrier. In particular, *E. coli* will rely on transport of the dissociated acid, while in *S. cerevisiae* an ABC transporter is the most efficient for efflux of carboxylic acids.

NExT-EMA was developed to perform NET-analysis on elementary flux modes and distinguish between thermodynamically feasible and infeasible solutions. The applicability of *NExT-EMA* to real world problems and advantages over existing methods was demonstrated by analysing adipic acid production in prokaryotic and eukaryotic models.

The analysis illustrates that high adipic acid yields in a biological system are theoretically achievable but significant differences exist between yields and thermodynamic feasibility of the different pathways. Examples are the infeasible inositol pathway and the partially infeasible lysine degradation in *E. coli*, as well as the infeasible route via hexa-2-enedioate of the 3-oxoadipate pathway in *S. cerevisiae*. Low pH capability influenced the thermodynamic equilibrium of the critical reaction in the inositol pathway in favour of product formation. In *S. cerevisiae* peroxisomal pH was found to have a crucial thermodynamic impinge on FA energy metabolism. Yields of FA based pathways appeared to be strongly dependent on the applied carbon-source and capability of acetyl-CoA recycling as a by-product of β-oxidation. In some cases, extreme differences between the organisms were found, in particular the 2-oxoadipate based routes delivered high yields in *E. coli* but featured the lowest in *S. cerevisiae*. In both organisms, the routes via 3-oxoadipate achieve the highest yields but are limited by a thermodynamic equilibrium on the substrate side. The only thermodynamically favoured routes are shikimate pathway based (dehydroshikimate, anthranilate and pHBA routes), which can achieve moderate to high yields. In a subsequent analysis aiming at commercial viability, where constraints were tightened, only the dehydroshikimate and isochorismate routes (in *E. coli*) were robust enough to deliver product at a high yield however at cost of low product titer. Other pathways could achieve higher titers, but at significantly lower yields. Still, publications on the respective pathways are far behind the determined maxima and achieve at best mediocre product titers and yields, hinting at kinetic and/or product export limitations.

Despite the theoretical nature of our analysis or perhaps exactly due to that, we were able to identify fundamental predetermined restrictions for adipic acid formation, providing a profound guidance for the development of worthwhile bio-production. The ability to

792 determine infeasible flux patterns will be pivotal for strain and process design, as tailored
793 strain construction should avoid infeasible flux distributions.
794

Acknowledgements

A program that aided to estimate some of the Gibbs energies of formation was kindly provided by Jason Jooste (AIBN). This research was supported by the NeCTAR Research Cloud and by QCIF (http://www.qcif.edu.au). The NeCTAR Research Cloud is a collaborative Australian research platform supported by the National Collaborative Research Infrastructure Strategy.

Funding: This study was partially funded by the Australian Research Council (DE120101549) and the Chilean National Commission for Scientific and Technological Research (CONICYT) by the Becas Chile post-doctorado program.

Figure legends

Abstract Graphic: Workflow of the analysis depicting how the respective parts are tied into the whole study.

Figure 1: Overview of metabolic routes to adipic acid and its precursors, muconic and glucaric acid. Each major pathway has several sub-routes that utilize alternative biochemical conversions, which can significantly influence titer, yield and productivity. Biochemical reduction of muconic and glucaric acid to adipic acid, has not yet been accomplished and is therefore indicated as hypothetical with dashed lines.

Figure 2: Logarithmic plot of internal and external adipic acid concentration ratios of the different studied transport mechanisms for *E. coli* (red) and *S. cerevisiae* (blue). The four different scenarios are passive diffusion of neutralized acid, proton anionic symporter, transport of dissociated acid with ATP usage, and ABC-transporter operating at equilibrium.

Figure 3: Maximum carbon yields for pathways to production of adipic acid (and its direct precursors) and impact of thermodynamics on these. Yields are given in % [C-mol/C-mol]. (A) *E. coli* and (B) *S. cerevisiae*.

Figure 4: Numbers of EFMs for studied pathways to production of adipic acid, total EFMs (blue), product forming EFMs (red) and thermodynamically feasible EFMs (green), in (A) *E. coli* and (B) *S. cerevisiae*.

Figure 5: Product vs. biomass carbon yield plots of feasible and infeasible modes from NExT-EMA analysis of certain shikimate pathway based routes in E. coli. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %. Color code: feasible EFMs under physiological conditions (blue), feasible EFMs under industrially relevant conditions (green), infeasible EFMs under physiological conditions (pink), infeasible EFMs under industrially relevant conditions (red). (A) Feasible EFMs of the dehydroshikimate route, (B) Infeasible EFMs of

the dehydroshikimate route, (C) Feasible EFMs of the isochorismate route, (D) Infeasible EFMs of isochorismate route.

Figure 6: Glycolysis based pathways for biological production of the principal adipic acid precursors muconic and glucaric acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

Figure 7: TCA cycle based pathways for biological production of adipic acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

Figure 8: Fatty acid metabolism based pathways for biological production of adipic acid. Different pathways are written in bold and are underlined. Rather than distinguishing between different routes to adipic acid this variant for adipic acid production is determined by the type of carbon-source used (not apparent from the figure) and the points of entry into the pathway cascade, which are written in bold. Dashed lines indicate multiple steps.

Figure 9: Flowchart for NET-analysis of elementary flux modes. EMA = elementary mode analysis; CCM = component contribution method; NET-analysis = network-embedded thermodynamic analysis; sgn = sign function: If EFMs are not already binary, reactions were constrained according to the direction (≙ sign) of the flux determined, for zero flux the reaction was considered unconstrained.

Appendix A

A.1 Handling of metabolic networks and assumptions for thermodynamic data

A.1.1 Principles, (biophysical) data and assumptions for construction of models

While glucose transport in *E. coli* and *S. cerevisiae* can be considered non-limiting under most circumstances [105], the exact uptake mechanism for FA into the cell is currently unclear. Given that the lipophilic character of the uncharged alkane residue dominates over the carboxylic acid function [106] transport should purely depend on the concentration gradient, oil/water interface and partitioning coefficient. Due to the lack of data we assumed free and non-limiting palmitate influx. *Ab initio* reaction directionalities were inferred from experimentally curated data (EcoCyc [107], SGD [82]). Water was not balanced (since water is necessarily ubiquitous in a biological system) for enumeration of elementary flux modes, in order to reduce the number of irrelevant modes.

However, water was naturally respected for thermodynamic analysis. Also, protons were balanced in all reactions for thermodynamics analysis (cf. supplementary files 6 & 7, "Models" sheet). Further, reactions that consume or produce CO_2, needed to be balanced with one equivalent of H_2O on the opposing side of the reaction to balance oxygen, as in the system the total CO_2 (co2tot) is distributed among the species HCO_3^-, CO_3^{2}, CO_2 and H_2CO_3 in aqueous phase [108]. Global carbon balancing reactions (*E. coli* R1 – R13 and *S. cerevisiae* R1 – R11) are irrelevant for NET-analysis and were therefore disregarded for thermodynamic analysis, as was the biomass equation (determination of Δ_rG not applicable, reaction assumed feasible). The Gibbs energies of all other remaining reactions were estimated and naturally taken into account for the thermodynamics analyses. The full networks are included as supplementary material (supplementary files 2 & 3 contain the models for enumeration of EFMs and supplementary files 6 & 7 contain the models for thermodynamic analysis) as example for use with *NExT-EMA*.

Thermodynamic data of compounds ($\Delta_f G^0$), including pseudoisomers, was obtained using the website version (http://equilibrator.weizmann.ac.il) of the eQuilibrator tool [109]. Experimental data [108, 110] was always preferred over component contribution method [66] derived values, if neither was available it was derived by means of group contribution method [111] principles. The different Gibbs energy of formation data sources were handled as described before [61], in order to avoid potential inconsistencies. Biophysical data (pH, ionic strength, redox potential, relative size) for the different compartments (extracellular, cytosol, mitochondrial matrix & intermembrane space, peroxisome) was compiled from different sources. In detail an ionic strength of 0.15 M was assumed for all compartments [57] in both organisms. Respective standard extracellular and cytosolic pH levels of 7.0 and 7.6 were used for *E. coli* [72]. For low pH conditions 5.9 (extracellular) and 6.1 (cytosol) were used [103]. For *S. cerevisiae* an extracellular pH of 6.7, a cytosolic pH of 7.2 and mitochondrial pH levels of 7.5 (matrix) and 6.4 (inter-membrane space) were used as standard conditions [71]. Outgoing from reported low extracellular, cytosolic pHs (3.5 and 4.4) [101, 102] the mitochondrial values were estimated (4.7 and 3.6) by assuming the same gradient as at standard conditions. As peroxisomal pH is highly disputed (latest theories indicate variability due to a Donnan equilibrium [112], suggesting a variable pH gradient [113]) two scenarios with the maximum (8.2) and minimum (5.8) [113] reported pHs were investigated. Redox potentials were estimated accordingly: based on the proton motive force and the cytosolic redox potential for *S. cerevisiae* of -286 mV [114] the membrane potential was related to the pH gradient as described previously [113], resulting in peroxisomal redox potentials of -345 mV and -203 mV respectively. The extracellular redox potential was estimated to be -196 mV, based on the value for the cytosol and an average trans-membrane potential of $\Delta\psi$ = 90 mV [115, 116]. Outgoing from this and an average trans-membrane potential of $\Delta\psi$ = 120 mV [117] the cytosolic redox potential of *E. coli* was accordingly estimated to be -316 mV. Mitochondrial (-296 mV) and inter-membrane

(-219 mV) redox potentials were adapted from literature [114] as done before [61, 67]. Volume of the peroxisomal compartment was adjusted depending on the pathway / carbon-source used: while under conditions of growth on glucose the peroxisomes typically only make up 1 to 2% of the total volume of the cells, it increases with FAs and certain other substrates, in some cases reaching up to 80% [118]. Further, when applying more accurate constraints through limited metabolite concentrations, the significance of the results is improved. Where available for *E. coli* and *S. cerevisiae* metabolism, maximum and minimum metabolite concentrations were applied. These were adopted from previous analyses [56, 63] or limited between 0.0001 mM and 10 mM [57, 61]. All biophysical data is also included in supplementary files 6 & 7.

A.1.2 Errors Gibbs energies estimated with *NExT-EMA*

NExT, and thus also *NExT-EMA*, does not consider errors in the $\Delta_f G^0$ estimates or the influence of ionic strength uncertainties nor the propagation into the final $\Delta_r G^0$ estimates. However, this analysis considers wide ranges of metabolite concentration, which have been previously shown to dwarf the uncertainty of the estimated $\Delta_r G$ values [61]. Similarly, errors in $\Delta_r G^{0'}$ due to ionic strength are small, a 0.1 M error will result in less than a 2 kJ/mol error in the $\Delta_r G^{0'}$. Furthermore, results have been carefully analysed, the cases where the extreme $\Delta_r G$ is less than the error of component contribution method or ionic strength have been tagged as borderline, in order to be considered with caution (see supplementary files 6 & 7, "infeasible_patterns" sheet).

A.2 Considerations, assumptions and calculations for tightened constraints

A.2.1 Definition of thresholds for yield, titer and rate

With industrial application in mind, the pathway networks were subject to analysis under tightened conditions. Metabolite constraints, in particular the minimum concentrations, were tightened based on performance criteria dictated by economic viability regarding yield, titer and rate. Specifically,

1. *Minimize substrate costs:* pathways were filtered based on a minimum product carbon yield; only EFMs that achieved a product carbon yield higher than 50% or 75% (two different scenarios) were subject to the tightened thermodynamic analysis.
2. *Minimize downstream purification costs:* a lower limit for the external product concentration (100 mM or 5 mM) was introduced. This requirement affects the minimal internal product concentration and hereby influences metabolite concentrations.
3. *Maximise productivity of capital infrastructure*: a minimum flux rate of 0.01 mol/(g_{CDW}×h) was assumed. This restriction constrains the minimum internal metabolites concentration as explained in A.2.2.

These industry requirements for yield, titer and rate were slightly adjusted from published values to suite the nature of the present study, based on the following rationales:

For a biotechnological production process, the yield is typically higher than 85% of the theoretical maximum [90] – as the present study compares different pathways (rather than different products) a constant threshold of 75% minimum total carbon yield was introduced in order to filter the pathways. Depending on the pathway this is in accordance with the reported industry requirement: if the Y_{max} of a certain pathway is lower than 88.2% the criterion is met. For the lower constant threshold of 50% minimum product carbon yield, pathways with a Y_{max} lower than 58.8% still meet the criterion.

In a biotechnological production process, titers need to be at least 50 g/L [91] – solubility of adipic acid in water is slightly lower, depending on the temperature (≈ 30 – 49 g/L), which limits the upper end. Further, the total concentration of all intracellular metabolites is commonly 300 mM [119], while the measured maximum for a single species is just below 100 mM. Thus, in order to avoid overcrowding of the intracellular volume, the maximum internal adipic acid concentration was limited to 100 mM (≈ 15 g/L). The minimum external product concentrations of 5 mM and 100 mM were chosen based on the outcome of the respective analyses, to achieve a meaningful clustering of the pathways and allow classification of the results.

Biotechnological production processes typically achieve rates around 2 g/(L×h) [91] – here a rate of 0.01 mol/(g_{CDW}×h) (≈ 2 g/(g_{CDW}×h)), was used. This can be achieved with at least 1 g/L biomass, which is on the lower end of typical large-scale fermentation processes.

A.2.2 Constraining internal metabolite concentrations

At low substrate concentration, [S] << K_m, the rate through an enzyme is dictated by $v = [E] \times (k_{cat}/K_m) \times [S]$. There are limits for how high the enzyme concentration ([E]) can be and how high the specificity of the enzyme (k_{cat}/K_m) can be. Accordingly, there is also a limit to how low the concentration [S] can be, when the rate must be at least 0.01 mol/g_{CDW}×h.

We will assume that no enzyme contributes more than 1% to the total proteome. This corresponds to the most abundant enzyme in *E. coli*, namely the acetyl-carrier protein (ACP), with up to 60.000 copies per cell, i.e., 1×10^{-19} mol/cell or 8.6×10^{-16} g/cell (when considering the M_w of ACP) [120]. With a minimum mass of a single *E. coli* cell of 1.5×10^{-13} g_{CDW}/cell [121] and a protein content of 0.5 g/g_{CDW} [122], this translate to an ACP fraction of ca. 1% of the whole proteome. A similar fraction is found for a fully induced LacZ [123]. Using an average M_w for enzymes of 50 kDa [124], the maximum theoretical enzyme concentration can be calculated to be $[E]_{max} = 1\times10^{-7}$ mol/g_{CDW}.

The 90th percentile of k_{cat}/K_m reported for all EC classes is approx. 1×10^{-7} 1/(s×M) [125]. If we use this as the maximum value, then $(k_{cat}/K_m)_{max} \times [E]_{max}$ becomes 1 L/(s×g_{CDW}) and the minimum concentration can be determined from $[S]_{min} = v_{min}/(1\ L/(s\times g_{CDW}) \times 3600\ s/h)$. With a required v_{min} of 0.01 mol/(g_{CDW}×h), the minimum substrate concentration of any reaction becomes 0.003 mM. Using different k_{cat}/K_m values for the six enzyme classes, the minimum substrate concentrations for each metabolite can be differentiated. For detailed step-by-step calculations and the resulting minimal metabolite concentrations for the individual scenarios, please refer to supplementary file 8.

A.3 Troubleshooting algorithm for identification of infeasibility-patterns

A workflow (cf. supplementary file 11) was developed to identify sets of directionalities (patterns) that cause a group of EFMs to be infeasible.

The algorithm consists of an iterative process, where first a reaction that, if removed would render the model thermodynamically feasible, is identified. Then it verifies that the directionality in the particular reaction is sufficient to render the model infeasible. These steps are repeated including further reactions until the smallest set of reactions, which is required to render the model infeasible is found.

First, feasible and infeasible EFMs were compared by means of the directionality of the reactions belonging to these two groups, i.e. the reactions that always had the same direction in all feasible or all infeasible EFMs were identified. By comparing these two groups of reactions, combinations of reactions (patterns) that potentially caused infeasibility could be identified. In case no clear differences were found between the directionality of feasible and infeasible EFMs, the constraints of the first infeasible EFM were used as input for the troubleshooting algorithm to identify the infeasibility-patterns.

The identified pattern, or first infeasible EFM constraints, were analysed with *NExT* using a novel troubleshooting algorithm. The algorithm identified a set of directionalities that under the specified conditions of metabolite concentrations deemed an EFM thermodynamically infeasible. We named this set of directionality-constrains *infeasible-pattern*. In detail, the first reaction/s of the pattern was/were identified by a thermodynamic feasibility check starting with a model without directionality constraints. Then directionality constraints were added incrementally one by one and the feasibility of the model was checked for each additional constraint: when an added constraint caused the model to become infeasible, the reaction was removed as constraint and selected as a candidate for an *infeasible-pattern*. After the first round with all constraints, the identified conflicting directions were re-evaluated, checking if

the directionalities in the particular reactions were sufficient to render the model infeasible. If the model was thermodynamically feasible a second round of testing of the constraints was performed. In the second round the model constraints considered initially were the ones identified in the first round, the remaining directionality constrains were added one by one. Analogue to the first round, more directionality constraints that would make the model infeasible were identified. The process was repeated until the full pattern of infeasibility constrains was identified, i.e. the selected directionality constraints generated a thermodynamically infeasible model.

Finally, the infeasible EFMs were crosschecked in order to verify that removing the identified pattern constraints would cause these previously infeasible EFMs to become feasible. If some EFMs remained infeasible after the removal of the pattern constraints, the process was repeated with the remaining infeasible EFMs, in order to identify additional *infeasible-patterns*.

Appendix B

Supporting Information

Supplementary file 1: Adipic acid export analysis: interactive excel file with detailed step-by-step calculations of the different product export scenarios in *E. coli* and *S. cerevisiae*.

Supplementary files 2 & 3: EMA networks: excel files containing the stoichiometric networks of *E. coli* and S. *cerevisiae* for Elementary Mode Analysis. Sheets in the excel files are sorted by pathway, with the last sheet containing additional information regarding the literature the pathways were compiled from, key intermediates in order to identify the pathways, as well as maximum product carbon yields for adipic, muconic and glucaric acid.

Supplementary files 4 & 5: Product carbon yield vs. biomass plots of feasible and infeasible modes from *NExT-EMA* analysis of *E. coli* and *S. cerevisiae*. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %.

Supplementary files 6 & 7: NET-analysis networks: Excel files containing the stoichiometric networks as well as physical data for compartments and metabolites of *E. coli* and *S. cerevisiae* for NET analysis and results of the different thermodynamic analyses, including detailed information on the infeasibility of the flux modes of each network (*infeasible-patterns*), ΔrG values of reactions at different pHs, and the tightened metabolite concentration ranges for the analysis regarding industrial viability.

Supplementary file 8: Excel files containing assumptions and calculations for determination of minimal metabolite concentrations based on enzyme kinetics.

Supplementary file 9: Details on feasibility of EFMs of the different pathways under constrained metabolite concentration ranges and different minima for production of the target compound (analysis for industrial viability).

Supplementary file 10: Plots of metabolite ranges in the feasible EFMs of the industrial viability analysis.

Supplementary file 11: Figure depicting workflow for identification of *infeasible-patterns*.

References

[1] Cavani, F., Centi, G., Perathoner, S., and Trifirò, F. (2009) *Sustainable Industrial Chemistry*, Wiley-VCH, Weinheim.

[2] Bart, J. C. J., and Cavallaro, S. (2015) Transiting from Adipic Acid to Bioadipic Acid. Part II. Biosynthetic Pathways, *Ind Eng Chem Res 54*, 567-576.

[3] Markets and Markets. (2016) Adipic Acid Market worth $7,539.2 Million by 2019. http://www.marketsandmarkets.com/PressReleases/adipic-acid.asp

[4] Grand View Research. (2016) Adipic Acid Market Analysis By Application (Nylon-6,6 Fiber, Nylon-6,6 Resin, Polyurethane, Adipate Ester) And Segment Forecasts To 2020. http://www.grandviewresearch.com/industry-analysis/adipic-acid-market

[5] IHS Chemical. (2012) Bio-Based Adipic Acid, In *Chemical Insight and Forecasting*, IHS Chemical.

[6] IHS Chemical. (2012) Chemical Economics Handbook: Adipic Acid, IHS Chemical.

[7] MCGroup. (2011) Adipic acid 2011 World Market Outlook and Forecast, Merchant Research & Consulting ltd. http://mcgroup.co.uk/researches/adipic-acid-adpa

[8] Bart, J. C. J., and Cavallaro, S. (2015) Transiting from Adipic Acid to Bioadipic Acid. 1, Petroleum-Based Processes, *Ind Eng Chem Res 54*, 1-46.

[9] Merchant Research & Consulting, L. (2017) Adipic Acid (ADPA): 2017 World Market Outlook and Forecast up to 2021, p 123.

[10] Musser, M. T. (2005) Adipic Acid, In *ULLMANN'S Encyclopedia of Industrial Chemistry* (Wiley-VCH, Ed.) 7 ed., Wiley-VCH, Weinheim.

[11] EPA. (1994) Background Report AP-42, Section 6.2, Adipic Acid Production, In *Compilation of Air Pollutant Emission Factors*, p 6, U.S. Environmental Protection Agency, OAQPS/TSD/EIB, Research Triangle Park.

[12] Alini, S., Basile, F., Blasioli, S., Rinaldi, C., and Vaccari, A. (2007) Development of new catalysts for N_2O-decomposition from adipic acid plant, *Appl Catal B Environ 70*, 323–329.

[13] Sato, K., Aoki, M., and Noyori, R. (1998) A "Green" route to adipic acid: direct oxidation of cyclohexenes with 30 percent hydrogen peroxide, *Science 281*, 1646-1647.

[14] Mainhardt, H., and Kruger, D. (2000) N_2O Emissions from Adipic acid and Nitric acid Production.

[15] Polen, T., Spelberg, M., and Bott, M. (2013) Toward biotechnological production of adipic acid and precursors from biorenewables, *J Biotechnol 167*, 75-84.

[16] Niu, W., Draths, K. M., and Frost, J. W. (2002) Benzene-free synthesis of adipic acid, *Biotechnol Prog 18*, 201-211.

[17] Vardon, D. R., Rorrer, N. A., Salvachua, D., Settle, A. E., Johnson, C. W., Menart, M. J., Cleveland, N. S., Ciesielski, P. N., Steirer, K. X., Dorgan, J. R., and Beckham, G. T. (2016) *cis,cis*-Muconic acid: separation and catalysis to bio-adipic acid for nylon-6,6 polymerization, *Green Chem*.

[18] Frost, J. W., and Draths, K. M. (1997) Bacterial cell transformation for production of *cis,cis*-muconic acid and catechol, Purdue Research Foundation, US.

[19] Frost, J. W., and Draths, K. M. (1996) Synthesis of adipic acid from Biomass-derived carbon sources, Purdue Research Foundation, US.

[20] Chen, Y., and Nielsen, J. (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks, *Curr Opin Biotechnol*.

[21] Deng, Y., Ma, L., and Mao, Y. (2016) Biological production of adipic acid from renewable substrates: Current and future methods, *Biochem Eng J 105, Part A*, 16-26.

[22] Xie, N.-Z., Liang, H., Huang, R.-B., and Xu, P. (2014) Biotechnological production of muconic acid: current status and future prospects, *Biotechnol Adv 32*, 615-622.
[23] Van Duuren, J. B. J. H., and Wittmann, C. (2014) First and Second Generation Production of Bio-Adipic Acid, In *Bioprocessing of Renewable Resources to Commodity Bioproducts*, pp 519-540, John Wiley & Sons, Inc.
[24] Lee, J.-H., and Wendisch, V. F. (2016) Biotechnological production of aromatic compounds of the extended shikimate pathway from renewable biomass, *J Biotech*.
[25] Leavitt, J. M., Wagner, J. M., Tu, C. C., Tong, A., Liu, Y., and Alper, H. S. (2017) Biosensor-Enabled Directed Evolution to Improve Muconic Acid Production in *Saccharomyces cerevisiae*, *Biotechnol J*, 1600687-n/a.
[26] Sun, X., Lin, Y., Huang, Q., Yuan, Q., and Yan, Y. (2013) A Novel Muconic Acid Biosynthetic Approach by Shunting Tryptophan Biosynthesis via Anthranilate, *Appl Environ Microbiol*.
[27] Weber, C., Bruckner, C., Weinreb, S., Lehr, C., Essl, C., and Boles, E. (2012) Biosynthesis of *cis,cis*-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*, *Appl Environ Microbiol 78*, 8421-8430.
[28] Curran, K. A., Leavitt, J. M., Karim, A. S., and Alper, H. S. (2013) Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*, *Metab Eng 15*, 55-66.
[29] Lin, Y., Sun, X., Yuan, Q., and Yan, Y. (2014) Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in *Escherichia coli*, *Metab Eng 23*, 62-69.
[30] Sengupta, S., Jonnalagadda, S., Goonewardena, L., and Juturu, V. (2015) Metabolic Engineering of a Novel Muconic Acid Biosynthesis Pathway via 4-Hydroxybenzoic Acid in *Escherichia coli*, *Appl Environ Microb 81*, 8037-8043.
[31] Vardon, D. R., Franden, M. A., Johnson, C. W., Karp, E. M., Guarnieri, M. T., Linger, J. G., Salm, M. J., Strathmann, T. J., and Beckham, G. T. (2015) Adipic acid production from lignin, *Energ Environ Sci 8*, 617-628.
[32] Yu, J.-L., Xia, X.-X., Zhong, J.-J., and Qian, Z.-G. (2014) Direct biosynthesis of adipic acid from a synthetic pathway in recombinant *Escherichia coli*, *Biotechnol Bioeng 111*, 2580-2586.
[33] Deng, Y., and Mao, Y. (2015) Production of adipic acid by the native-occurring pathway in *Thermobifida fusca* B6, *J Appl Microbiol 119*, 1057-1063.
[34] Clomburg, J. M., Blankschien, M. D., Vick, J. E., Chou, A., Kim, S., and Gonzalez, R. (2015) Integrated engineering of β-oxidation reversal and ω-oxidation pathways for the synthesis of medium chain ω-functionalized carboxylic acids, *Metab Eng 28*, 202-212.
[35] Reizman, I. M. B., Stenger, A. R., Reisch, C. R., Gupta, A., Connors, N. C., and Prather, K. L. J. (2015) Improvement of glucaric acid production in *E. coli* via dynamic control of metabolic fluxes, *Metab Eng Commun 2*, 109-116.
[36] Babu, T., Yun, E. J., Kim, S., Kim, D. H., Liu, K. H., Kim, S. R., and Kim, K. H. (2015) Engineering Escherichia coli for the production of adipic acid through the reversed β-oxidation pathway, *Process Biochem*.
[37] Sun, X., Lin, Y., Yuan, Q., and Yan, Y. (2014) Biological production of muconic acid via a prokaryotic 2,3-dihydroxybenzoic acid decarboxylase, *ChemSusChem 7*, 2478-2481.
[38] Horwitz, Andrew A., Walter, Jessica M., Schubert, Max G., Kung, Stephanie H., Hawkins, K., Platt, Darren M., Hernday, Aaron D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, Sunil S., and Newman, Jack D. (2015) Efficient Multiplexed

Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas, *Cell Syst 1*, 88-96.

[39] Raemakers-Franken, P. C., Schürmann, M., Trefzer, A. C., and De Wildeman, S. M. A. (2012) Preparation of adipic acid, DSM IP Assets B.V., US.

[40] Picataggio, S., and Beardslee, T. (2012) Biological methods for preparing adipic acid, Verdezyne, Inc., US.

[41] Burgard, A. P., Pharkya, P., and Osterhout, R. E. (2012) Microorganisms for the production of adipic acid and other compounds, Genomatica, Inc., US.

[42] Burgard, A. P., Pharkya, P., and Osterhout, R. E. (2010) Microorganisms for the production of adipic acid and other compounds, Genomatica, Inc., US.

[43] Boussie, T. R., Dias, E. L., Fresco, Z. M., and Murphy, V. J. (2010) Production of adipic acid and derivates from carbohydrate-containing materials, Rennovia, Inc., US.

[44] Guzman, D. d. (2010) Bio-adipic acid prepares for entry, *ICIS Chemical Business 278*, 22.

[45] Karlsson, E., Mapelli, V., and Olsson, L. (2017) Adipic acid tolerance screening for potential adipic acid production hosts, *Microb Cell Fac 16*, 20.

[46] Vickers, C. E., Klein-Marcuschamer, D., and Krömer, J. O. (2012) Examining the feasibility of bulk commodity production in *Escherichia coli*, *Biotechnol Lett 34*, 585-596.

[47] Schuster, S., and Hilgetag, C. (1994) On Elementary Flux Modes in Biochemical Reaction Systems at Steady State, *J Biol Systems 2*, 165-182.

[48] Krömer, J. O., Wittmann, C., Schroder, H., and Heinzle, E. (2006) Metabolic pathway analysis for rational design of *L*-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*, *Metab Eng 8*, 353-369.

[49] Gruchattka, E., and Kayser, O. (2015) *In Vivo* Validation of *In Silico* Predicted Metabolic Engineering Strategies in Yeast: Disruption of α-Ketoglutarate Dehydrogenase and Expression of ATP-Citrate Lyase for Terpenoid Production, *PLoS ONE 10*, e0144981.

[50] Chen, N., Du, J., Liu, H., and Xu, Q. (2009) Elementary mode analysis and metabolic flux analysis of L-glutamate biosynthesis by *Corynebacterium glutamicum*, *Ann Microbiol 59*, 317-322.

[51] Li, S., Huang, D., Li, Y., Wen, J., and Jia, X. (2012) Rational improvement of the engineered isobutanol-producing *Bacillus subtilis* by elementary mode analysis, *Microb Cell Fact 11*, 1-12.

[52] Chen, Z., Liu, H., Zhang, J., and Liu, D. (2010) Elementary mode analysis for the rational design of efficient succinate conversion from glycerol by *Escherichia coli*, *J Biomed Biotechnol 2010*, 518743.

[53] Diniz, S. C., Voss, I., and Steinbuchel, A. (2006) Optimization of cyanophycin production in recombinant strains of *Pseudomonas putida* and *Ralstonia eutropha* employing elementary mode analysis and statistical experimental design, *Biotechnol Bioeng 93*, 698-717.

[54] Averesch, N. J. H. (2016) Engineering yeast shikimate pathway towards production of aromatics: rational design of a chassis cell using systems and synthetic biology, In *School of Chemical Engineering*, The University of Queensland, UQ eSpace.

[55] Fleming, R. M., Thiele, I., and Nasheuer, H. P. (2009) Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*, *Biophys Chem 145*, 47-56.

[56] Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007) Thermodynamics-based metabolic flux analysis, *Biophys J 92*, 1792-1805.

[57] Kümmel, A., Panke, S., and Heinemann, M. (2006) Systematic assignment of thermodynamic constraints in metabolic network models, *BMC Bioinformatics 7*, 512.
[58] Mavrovouniotis, M. L. (1993) Identification of localized and distributed bottlenecks in metabolic pathways, *Proc Int Conf Intell Syst Mol Biol 1*, 275-283.
[59] Ataman, M., and Hatzimanikatis, V. (2015) Heading in the right direction: thermodynamics-based network analysis and pathway engineering, *Curr Opin Biotechnol 36*, 176-182.
[60] Kümmel, A., Panke, S., and Heinemann, M. (2006) Putative regulatory sites unraveled by network‐embedded thermodynamic analysis of metabolome data, *Molecular Systems Biology 2*.
[61] Martínez, V. S., Quek, L.-E., and Nielsen, Lars K. (2014) Network Thermodynamic Curation of Human and Yeast Genome-Scale Metabolic Models, *Biophys J 107*, 493-503.
[62] Boghigian, B. A., Shi, H., Lee, K., and Pfeifer, B. A. (2010) Utilizing elementary mode analysis, pathway thermodynamics, and a genetic algorithm for metabolic flux determination and optimal metabolic network design, *BMC Syst Biol 4*, 49.
[63] Jol, S. J., Kummel, A., Terzer, M., Stelling, J., and Heinemann, M. (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes, *PLoS Comput Biol 8*, e1002415.
[64] Gerstl, M. P., Ruckerbauer, D. E., Mattanovich, D., Jungreuthmayer, C., and Zanghellini, J. (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks, *Sci Rep 5*, 8930.
[65] Peres, S., Jolicœur, M., Moulin, C., Dague, P., and Schuster, S. (2017) How important is thermodynamics for identifying elementary flux modes?, *PLOS ONE 12*, e0171440.
[66] Noor, E., Haraldsdóttir, H. S., Milo, R., and Fleming, R. M. T. (2013) Consistent Estimation of Gibbs Energy Using Component Contributions, *PLoS Comput Biol 9*, e1003098.
[67] Martínez, V. S., and Nielsen, L. K. (2014) NExT: Integration of Thermodynamic Constraints and Metabolomics Data into a Metabolic Network, In *Metabolic Flux Analysis* (Krömer, J. O., Nielsen, L. K., and Blank, L. M., Eds.), pp 65-78, Springer New York.
[68] Maris, A. J. A. v., Konings, W. N., Dijken, J. P. v., and Pronk, J. T. (2004) Microbial export of lactic and 3-hydroxypropanoic acid: implications for industrial fermentation processes, *Metab Eng 6*, 245-255.
[69] Stephanopoulos, G. N., Aristidou, A. A., and Nielsen, J. (1998) *Metabolic Engineering: Principles and Methodologies*, Academic Press, San Diego.
[70] Urs von Stockar, L. A. M. v. d. W. (2013) *Biothermodynamics: The Role of Thermodynamics in Biochemical Engineering*, EPFL Press.
[71] Orij, R., Postmus, J., Ter Beek, A., Brul, S., and Smits, G. J. (2009) *In vivo* measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in *Saccharomyces cerevisiae* reveals a relation between intracellular pH and growth, *Microbiology 155*, 268-278.
[72] Zilberstein, D., Agmon, V., Schuldiner, S., and Padan, E. (1984) *Escherichia coli* intracellular pH, membrane potential, and cell growth, *J Bacteriol 158*, 246-252.
[73] Chen, J., Zhu, X., Tan, Z., Xu, H., Tang, J., Xiao, D., and Zhang, X. (2013) Activating C4-dicarboxylate transporters DcuB and DcuC for improving succinate production, *Appl Microbiol Biot 98*, 2197-2205.
[74] Beauprez, J. J., Foulquie-Moreno, M. R., Maertens, J., van Horen, E., Bekers, K., Baart, G. J. E., Cunin, R. M., Charlier, D., Heijnen, J. J., and Soetaert, W. (2011) Influence

of C4-dicarboxylic acid transporters on succinate production, *Green Chem 13*, 2179-2186.

[75] Moussatova, A., Kandt, C., O'Mara, M. L., and Tieleman, D. P. (2008) ATP-binding cassette transporters in *Escherichia coli*, *BBA-Biomembranes 1778*, 1757-1771.

[76] Casal, M., Paiva, S., Queirós, O., and Soares-Silva, I. (2008) Transport of carboxylic acids in yeasts, *FEMS Microbiol Rev 32*, 974-994.

[77] Taymaz-Nikerel, H., Jamalzadeh, E., Borujeni, A. E., Verheijen, P. J. T., Gulik, W. M. v., and Heijnen, J. J. (2013) A Thermodynamic Analysis Of Dicarboxylic Acid Production In Microorganisms, In *Biothermodynamics: The Role of Thermodynamics in Biochemical Engineering* (Stockar, U. v., Ed.), pp 547-579, EFPL Press, Lausanne Switzerland.

[78] Draths, K. M., and Frost, J. W. (1994) Environmentally Compatible Synthesis of Adipic Acid from D-Glucose, *J Am Chem Soc 116*, 399-400.

[79] Hardie, D. G., and Pan, D. A. (2002) Regulation of fatty acid synthesis and oxidation by the AMP-activated protein kinase, *Biochem Soc Trans 30*, 1064-1070.

[80] van Roermund, C. W. T., de Jong, M., IJlst, L., van Marle, J., Dansen, T. B., Wanders, R. J. A., and Waterham, H. R. (2004) The peroxisomal lumen in *Saccharomyces cerevisiae* is alkaline, *J Cell Sci 117*, 4231-4237.

[81] Rottensteiner, H., and Theodoulou, F. L. (2006) The ins and outs of peroxisomes: Co-ordination of membrane transport and peroxisomal metabolism, *BBA-Mol Cell Res 1763*, 1527-1540.

[82] Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Simison, M., Weng, S., and Wong, E. D. (2012) Saccharomyces Genome Database: the genomics resource of budding yeast, *Nucleic Acids Res 40*, D700-705.

[83] Moon, T. S., Yoon, S. H., Lanza, A. M., Roy-Mayhew, J. D., and Prather, K. L. (2009) Production of glucaric acid from a synthetic pathway in recombinant *Escherichia coli*, *Appl Environ Microbiol 75*, 589-595.

[84] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Res 27*, 29-34.

[85] 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, *Biotechnol Bioeng 106*, 462-473.

[86] Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007) Thermodynamics-Based Metabolic Flux Analysis, *Biophys J 92*, 1792-1805.

[87] Schuster, S., Dandekar, T., and Fell, D. A. (1999) Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering, *Trends Biotechnol 17*, 53-60.

[88] Gerstl, M. P., Jungreuthmayer, C., and Zanghellini, J. (2015) tEFMA: computing thermodynamically feasible elementary flux modes in metabolic networks, *Bioinformatics*.

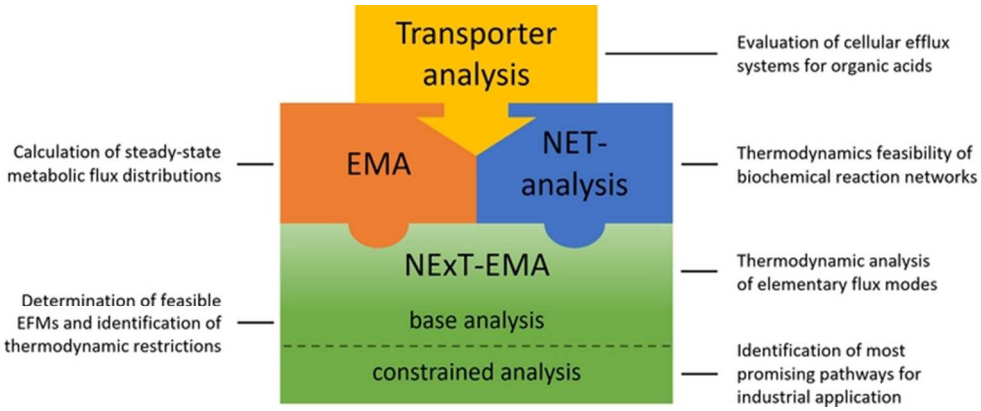
[89] van Klinken, J. B., and Willems van Dijk, K. (2015) FluxModeCalculator: an efficient tool for large-scale flux mode computation, *Bioinformatics*.

[90] Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced biofuels, *Nature 488*, 320-328.

[91] Woodley, J. M. (2017) Bioprocess intensification for the effective production of chemical products, *Comput Chem Eng 105*, 297-307.

[92] Zhang, H., Pereira, B., Li, Z., and Stephanopoulos, G. (2015) Engineering *Escherichia coli* coculture systems for the production of biochemical products, *Proc Natl Acad Sci USA 112*, 8266-8271.
[93] Ferguson, S. J. (2010) ATP synthase: From sequence to ring size to the P/O ratio, *Proc Natl Acad Sci USA 107*, 16755-16756.
[94] Steigmiller, S., Turina, P., and Graber, P. (2008) The thermodynamic H+/ATP ratios of the H+-ATPsynthases from chloroplasts and *Escherichia coli*, *Proc Natl Acad Sci USA 105*, 3745-3750.
[95] Lecchi, S., and Slayman, C. W. (2008) Plasma-Membrane H+-ATPase From Yeast, In *Protein Science Encyclopedia*, Wiley-VCH Verlag GmbH & Co. KGaA.
[96] Zhang, H., Li, Z., Pereira, B., and Stephanopoulos, G. (2015) Engineering *E. coli-E. coli* cocultures for production of muconic acid from glycerol, *Microb Cell Fac 14*, 134.
[97] Schrumpf, B., Schwarzer, A., Kalinowski, J., Pühler, A., Eggeling, L., and Sahm, H. (1991) A functionally split pathway for lysine synthesis in *Corynebacterium glutamicium*, *J Bacteriol 173*, 4510-4516.
[98] Mishra, P., Park, G.-Y., Lakshmanan, M., Lee, H.-S., Lee, H., Chang, M. W., Ching, C. B., Ahn, J., and Lee, D.-Y. (2016) Genome-scale metabolic modeling and in silico analysis of lipid accumulating yeast *Candida tropicalis* for dicarboxylic acid production, *Biotechnol Bioeng 113*, 1993-2004.
[99] Averesch, N. J. H., and Krömer, J. O. (2014) Tailoring strain construction strategies for muconic acid production in *S. cerevisiae* and *E. coli*, *Metab Eng Commun 1*, 19-28.
[100] Hiltunen, J. K., Mursula, A. M., Rottensteiner, H., Wierenga, R. K., Kastaniotis, A. J., and Gurvitz, A. (2003) The biochemistry of peroxisomal β-oxidation in the yeast *Saccharomyces cerevisiae*, *FEMS Microbiol Rev 27*, 35-64.
[101] Guldfeldt, L. U., and Arneborg, N. (1998) Measurement of the effects of acetic acid and extracellular pH on intracellular pH of nonfermenting, individual *Saccharomyces cerevisiae* cells by fluorescence microscopy, *Appl Environ Microbiol 64*, 530-534.
[102] Imai, T., and Ohno, T. (1995) The relationship between viability and intracellular pH in the yeast *Saccharomyces cerevisiae*, *Appl Environ Microbiol 61*, 3604-3608.
[103] Diez-Gonzalez, F., and Russell, J. B. (1997) The ability of *Escherichia coli* O157:H7 to decrease its intracellular pH and resist the toxicity of acetic acid, *Microbiology 143 (Pt 4)*, 1175-1180.
[104] Terzer, M., and Stelling, J. (2008) Large-scale computation of elementary flux modes with bit pattern trees, *Bioinformatics 24*, 2229-2235.
[105] Bisson, L. F., and Fraenkel, D. G. (1984) Expression of kinase-dependent glucose uptake in *Saccharomyces cerevisiae*, *J Bacteriol 159*, 1013-1017.
[106] Hamilton, J. A. (1998) Fatty acid transport: difficult or easy?, *J Lipid Res 39*, 467-481.
[107] Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., Latendresse, M., Muniz-Rascado, L., Ong, Q., Paley, S., Schroder, I., Shearer, A. G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R. P., Paulsen, I., and Karp, P. D. (2013) EcoCyc: fusing model organism databases with systems biology, *Nucleic Acids Res 41*, D605-612.
[108] Alberty, R. A. (2003) *Thermodynamics of Biochemical Reactions*, Wiley-VCH, Hoboken.
[109] Flamholz, A., Noor, E., Bar-Even, A., and Milo, R. (2012) eQuilibrator—the biochemical thermodynamics calculator, *Nucleic Acids Research 40*, D770-D775.
[110] Alberty, R. A. (2006) Appendix 2: Tables of Transformed Thermodynamic Properties, In *Biochemical Thermodynamics*, pp 409-424, John Wiley & Sons, Inc.

[111] Jankowski, M. D., Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks, *Biophys J 95*, 1487-1499.
[112] Antonenkov, V. D., and Hiltunen, J. K. (2012) Transfer of metabolites across the peroxisomal membrane, *BBA-Mol Basis Dis 1822*, 1374-1386.
[113] Haraldsdóttir, H. S., Thiele, I., and Fleming, R. M. T. (2012) Quantitative Assignment of Reaction Directionality in a Multicompartmental Human Metabolic Reconstruction, *Biophys J 102*, 1703-1711.
[114] Hu, J., Dong, L., and Outten, C. E. (2008) The Redox Environment in the Mitochondrial Intermembrane Space Is Maintained Separately from the Cytosol and Matrix, *J Biol Chem 283*, 29126-29134.
[115] Vacata, V., Kotyk, A., and Sigler, K. (1981) Membrane potential in yeast cells measured by direct and indirect methods, *BBA-Biomembranes 643*, 265-268.
[116] Lichtenberg, H. C., Giebeler, H., and Höfer, M. (1988) Measurements of electrical potential differences across yeast plasma membranes with microelectrodes are consistent with values from steady-state distribution of tetraphenylphosphonium inPichia humboldtii, *J Membrane Biol 103*, 255-261.
[117] Cell Biology by the Numbers. (2016) What is the electric potential difference across membranes? http://book.bionumbers.org/what-is-the-electric-potential-difference-across-membranes/
[118] Veenhuis, M., Salomons, F. A., and Van Der Klei, I. J. (2000) Peroxisome biogenesis and degradation in yeast: A structure/function analysis, *Microsc Res Techniq 51*, 584-600.
[119] Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*, *Nat Chem Biol 5*, 593-599.
[120] Boom, T. V., and Cronan, J. E. (1989) Genetics and Regulation of Bacterial Lipid Metabolism, *Annu Rev Microbiol 43*, 317-343.
[121] Neidhardt, F. C., and Curtiss, R. (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*, Vol. 1, ASM Press, Washington, D.C.
[122] Ertugay, N., and Hamamci, H. (1997) Continuous cultivation of bakers' yeast: Change in cell composition at different dilution rates and effect of heat stress on trehalose level, *Folia Microbiol 42*, 463-467.
[123] Tomer, K., Erez, D., and Uri, A. (2007) Cost–benefit theory and optimal design of gene regulation functions, *Phys Biol 4*, 229.
[124] Carroll, K. M., Simpson, D. M., Eyers, C. E., Knight, C. G., Brownridge, P., Dunn, W. B., Winder, C. L., Lanthaler, K., Pir, P., Malys, N., Kell, D. B., Oliver, S. G., Gaskell, S. J., and Beynon, R. J. (2011) Absolute quantification of the glycolytic pathway in yeast: deployment of a complete QconCAT approach, *Mol Cell Proteomics 10*.
[125] Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters, *Biochemistry 50*, 4402-4410.



Abstract Graphic / Table of Contents Graphic: Workflow of the analysis depicting how the respective parts are tied into the whole study.

63x25mm (300 x 300 DPI)

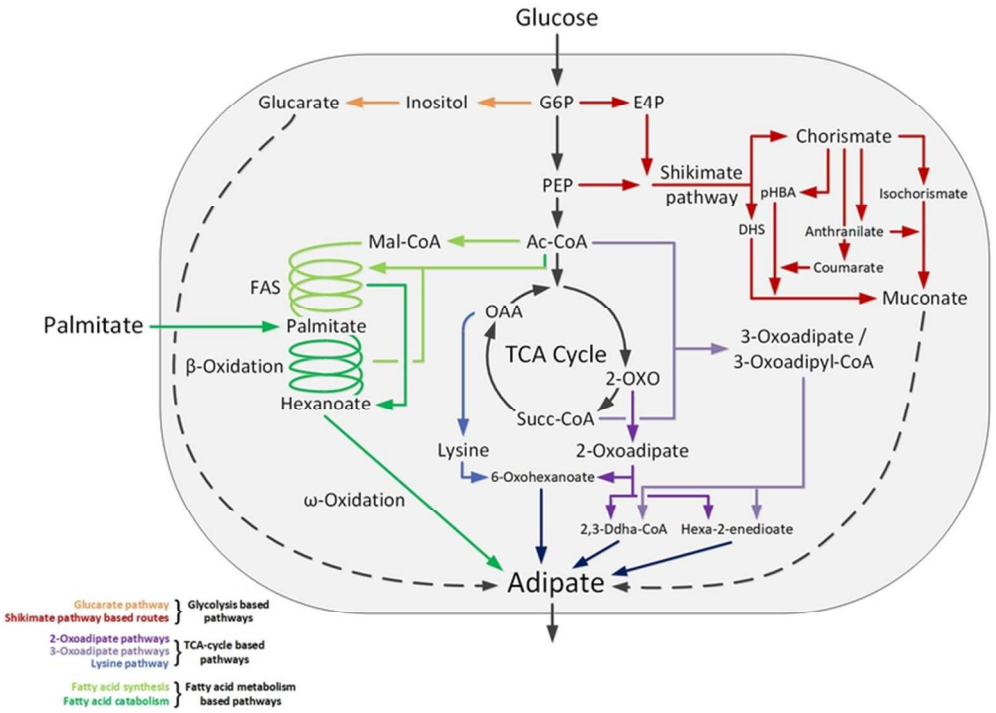


Figure 1: Overview of metabolic routes to adipic acid and its precursors, muconic and glucaric acid. Each major pathway has several sub-routes that utilize alternative biochemical conversions, which can significantly influence titer, yield and productivity. Biochemical reduction of muconic and glucaric acid to adipic acid, has not yet been accomplished and is therefore indicated as hypothetical with dashed lines.

69x49mm (300 x 300 DPI)

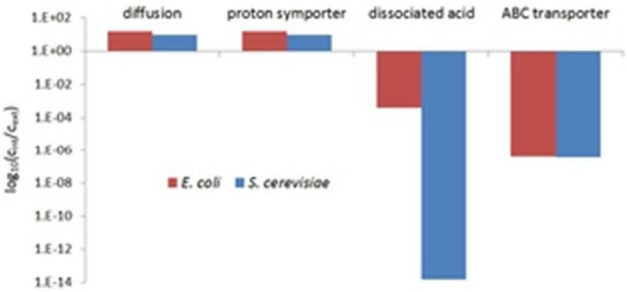


Figure 2: Logarithmic plot of internal and external adipic acid concentration ratios of the different studied transport mechanisms for *E. coli* (red) and *S. cerevisiae* (blue). The four different scenarios are passive diffusion of neutralized acid, proton anionic symporter, transport of dissociated acid with ATP usage, and ABC-transporter operating at equilibrium.

26x12mm (300 x 300 DPI)

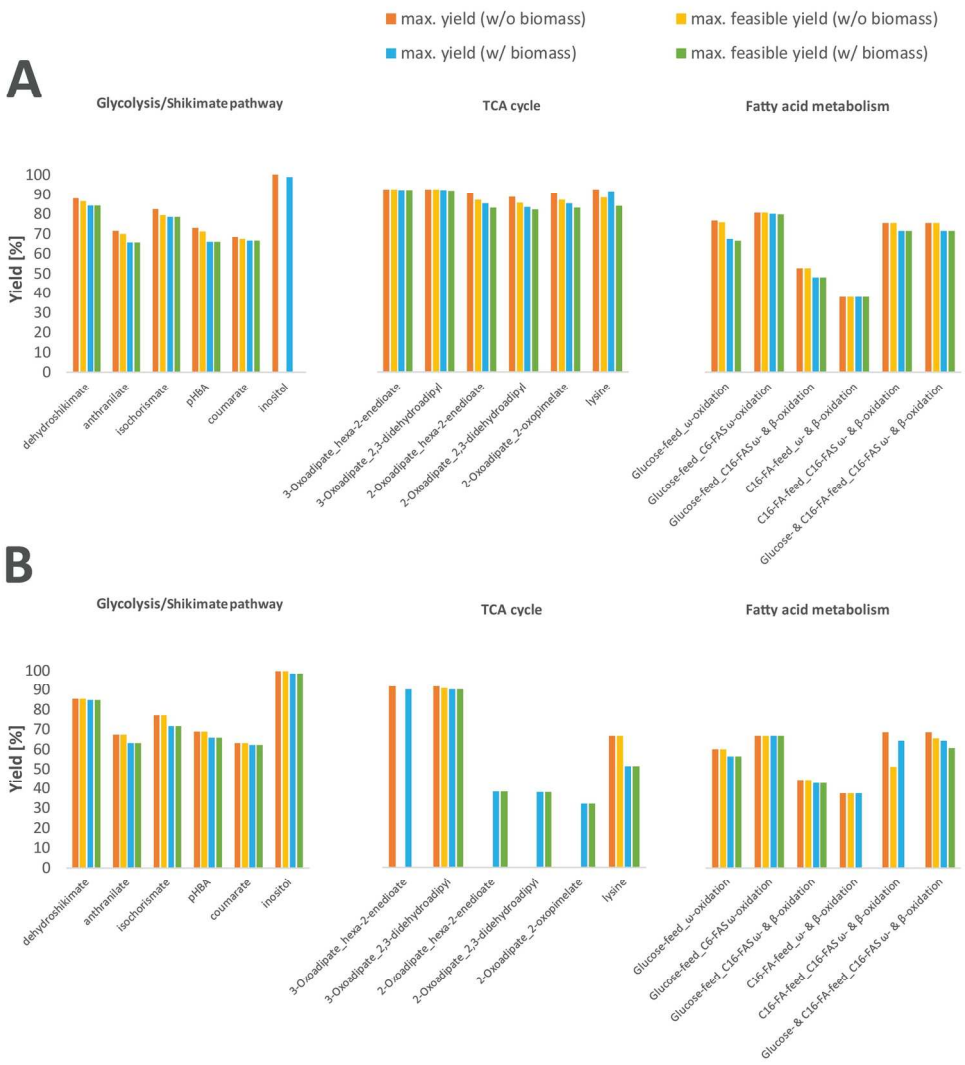


Figure 3: Maximum carbon yields for pathways to production of adipic acid (and its direct precursors) and impact of thermodynamics on these. Yields are given in % [C-mol/C-mol]. (A) *E. coli* and (B) *S. cerevisiae*.

171x185mm (300 x 300 DPI)

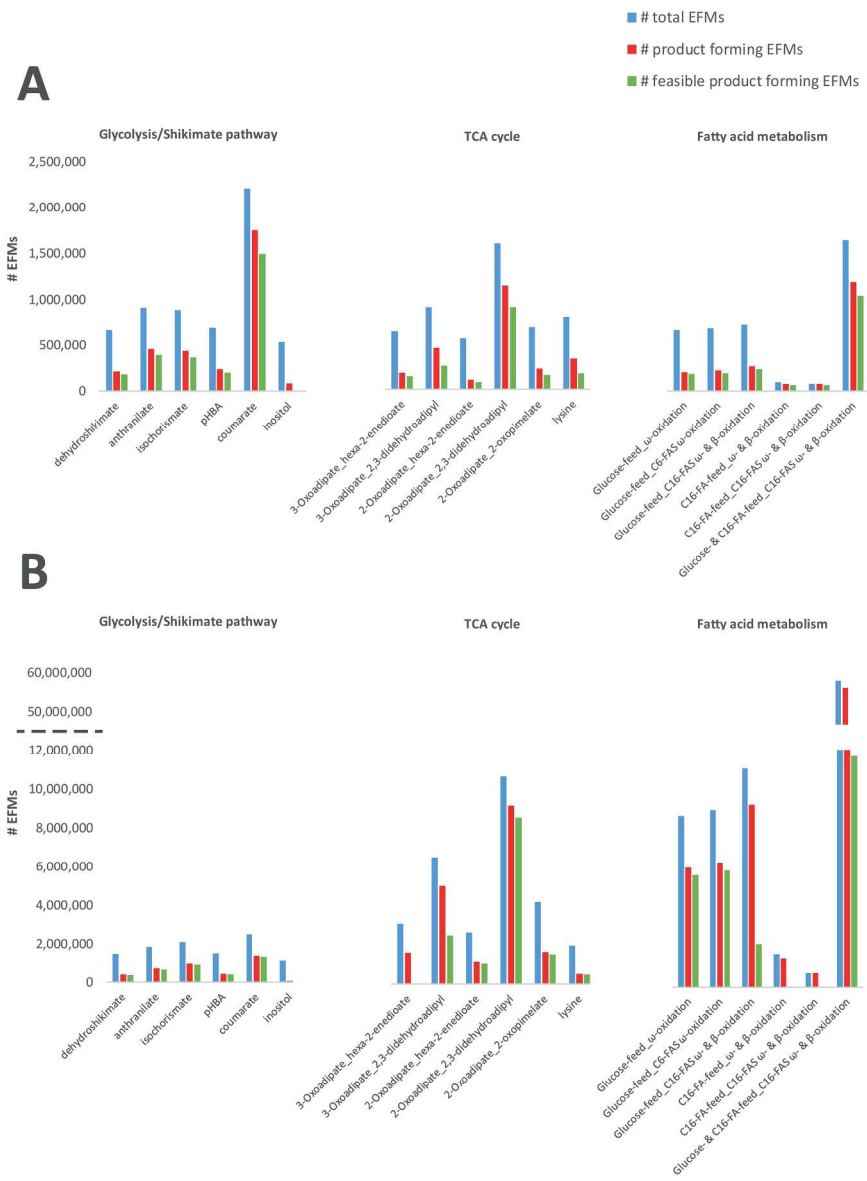


Figure 4: Numbers of EFMs for studied pathways to produce adipic acid, total EFMs (blue), product forming EFMs (red) and thermodynamically feasible EFMs (green), in (A) *E. coli* and (B) *S. cerevisiae*.

209x276mm (300 x 300 DPI)

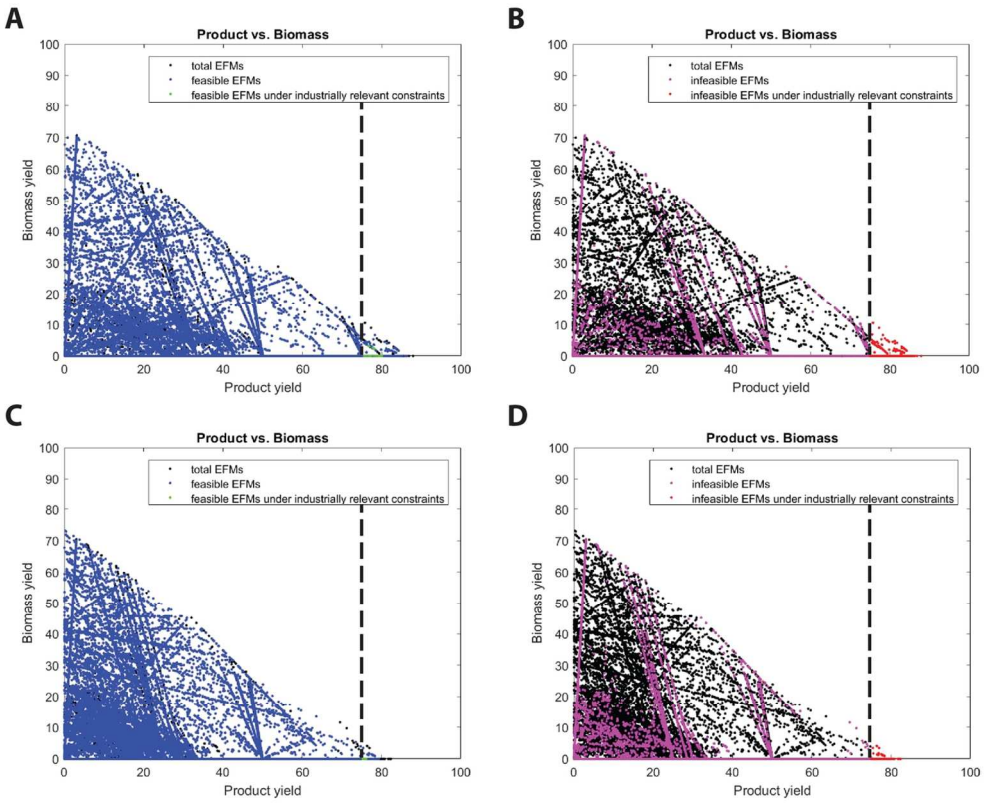


Figure 5: Product vs. biomass carbon yield plots of feasible and infeasible modes from NExT-EMA analysis of certain shikimate pathway based routes in *E. coli*. Each point in a chart corresponds to the specific product and biomass yield of the respective elementary flux mode. Yields are carbon yields in %. Color code: feasible EFMs under physiological conditions (blue), feasible EFMs under industrially relevant conditions (green), infeasible EFMs under physiological conditions (pink), infeasible EFMs under industrially relevant conditions (red). (A) Feasible EFMs of the dehydroshikimate route, (B) Infeasible EFMs of the dehydroshikimate route, (C) Feasible EFMs of the isochorismate route, (D) Infeasible EFMs of isochorismate route.

127x101mm (300 x 300 DPI)

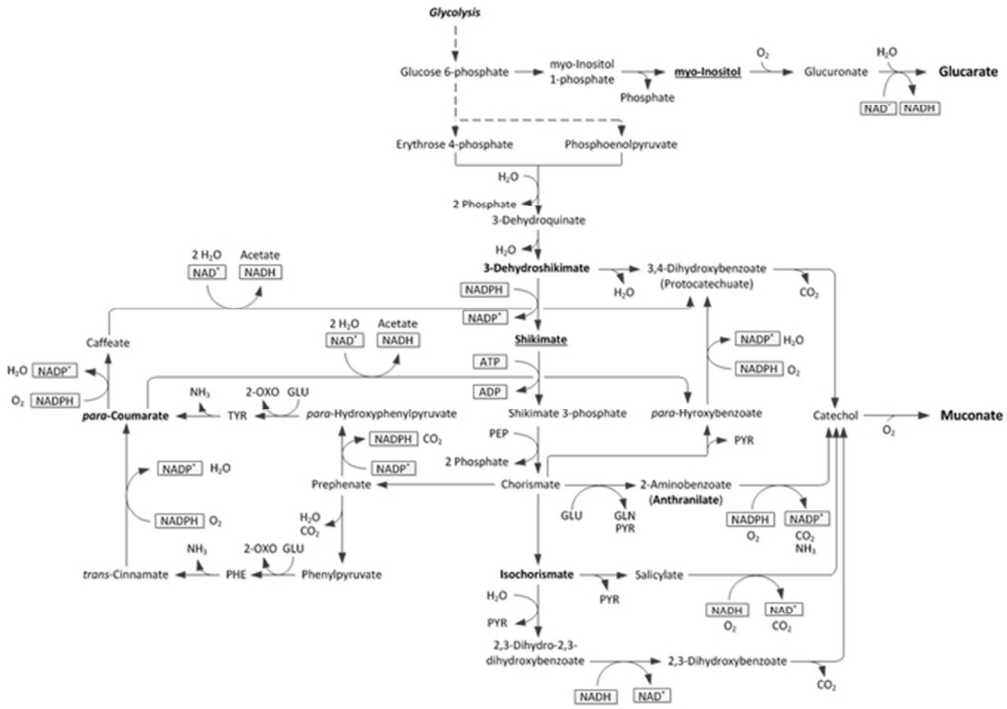


Figure 6: Glycolysis based pathways for biological production of the adipic acid precursors muconic and glucaric acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

57x40mm (300 x 300 DPI)

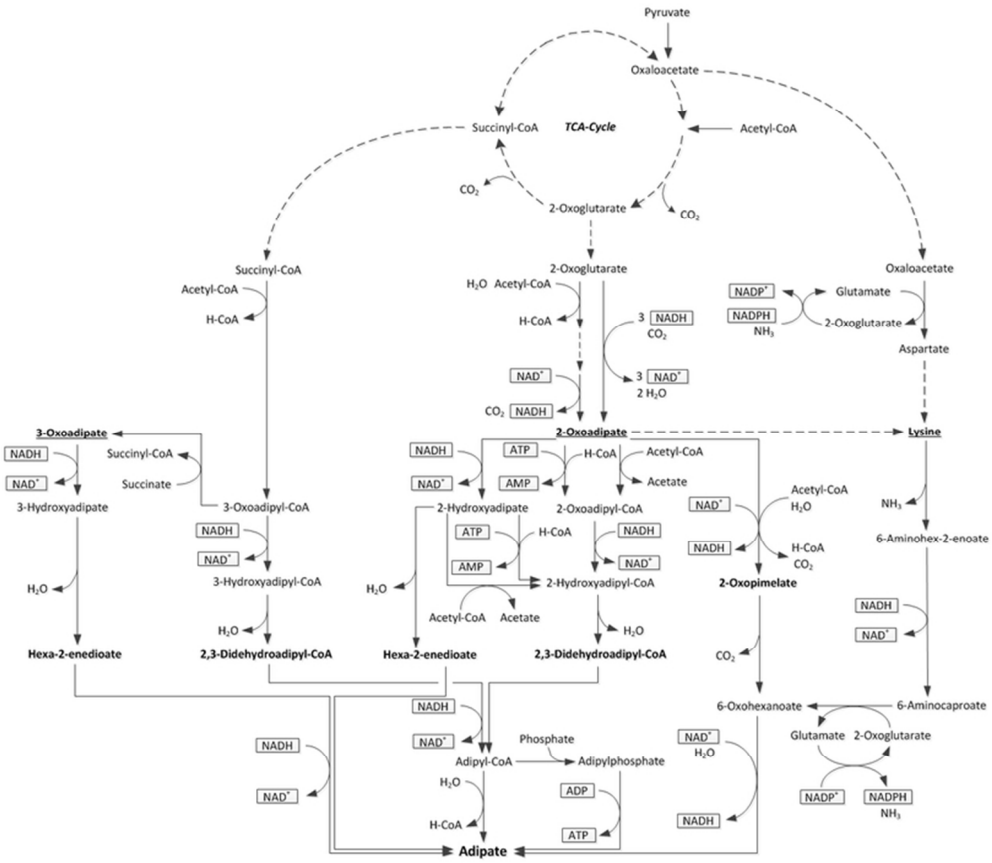


Figure 7: TCA cycle based pathways for biological production of adipic acid. Intermediates that define a pathway are written in bold and are underlined; central intermediates of a separate route are written in bold only. Dashed lines indicate multiple steps.

69x59mm (300 x 300 DPI)

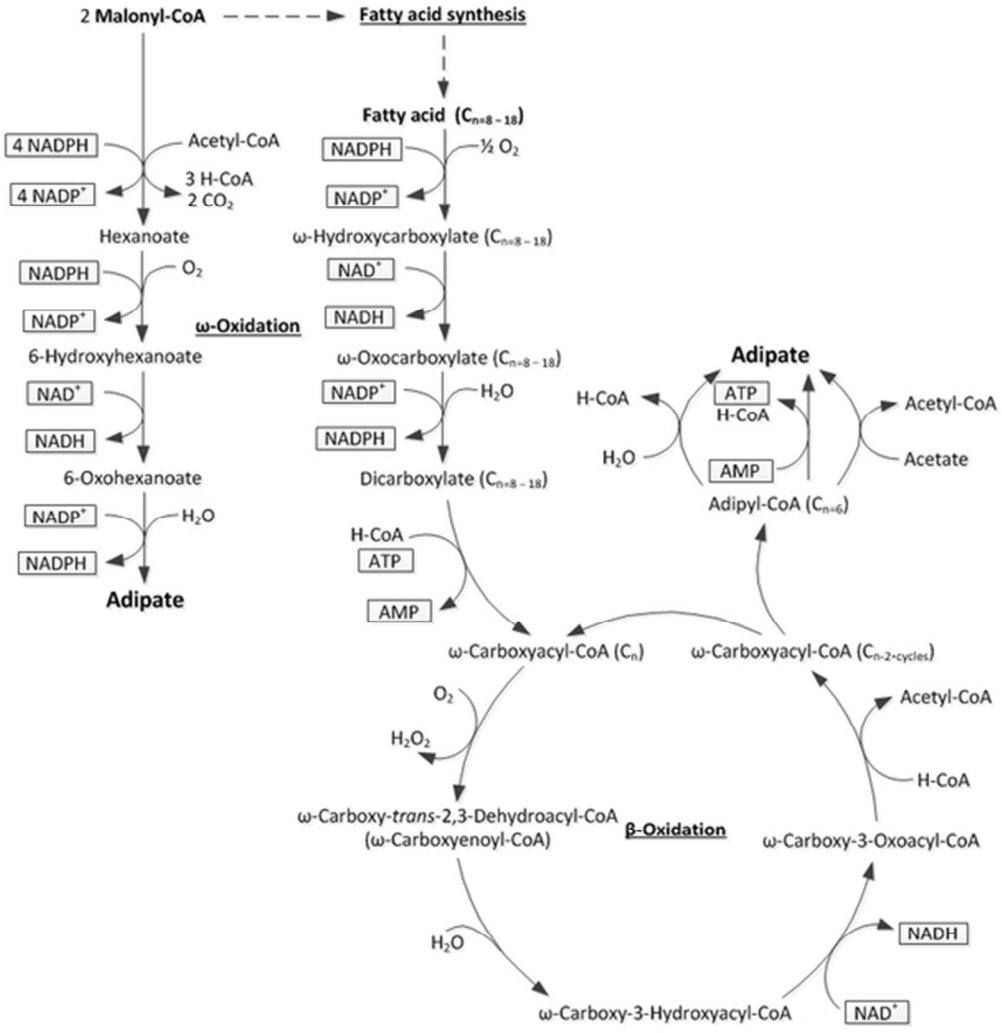


Figure 8: Fatty acid metabolism based pathways for biological production of adipic acid. Different pathways are written in bold and are underlined. Rather than distinguishing between different routes to adipic acid this variant for adipic acid production is determined by the type of carbon-source used (not apparent from the figure) and the points of entry into the pathway cascade, which are indicated in bold. Dashed lines indicate multiple steps.

51x53mm (300 x 300 DPI)

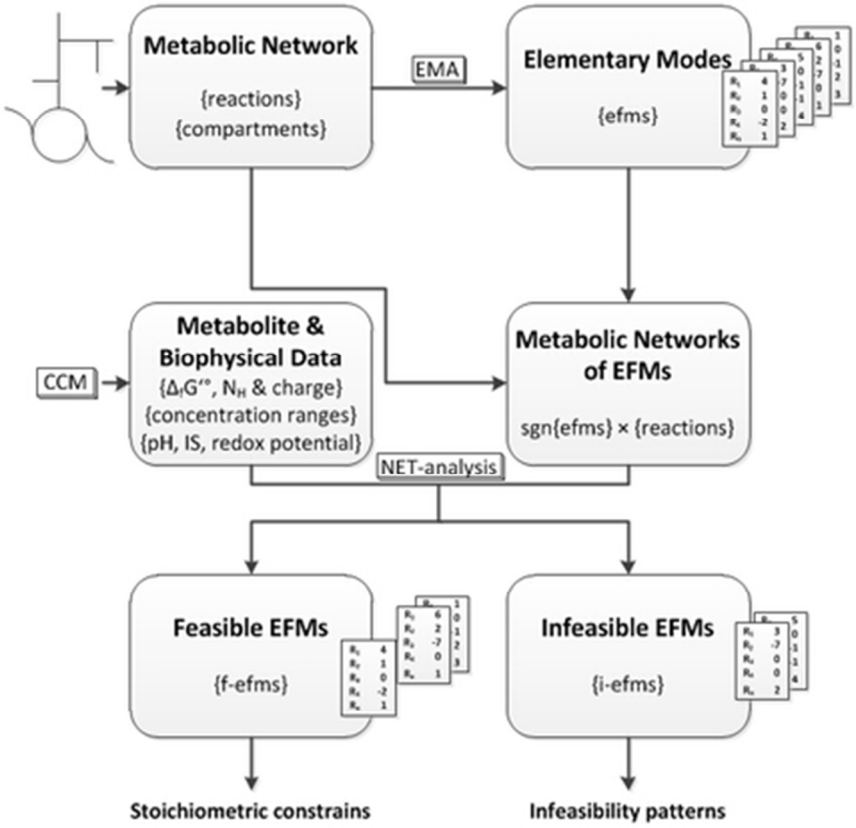


Figure 9: Flowchart for NET-analysis of elementary flux modes. EMA = elementary mode analysis; NET-analysis = network-embedded thermodynamic analysis; IS = ionic strength; sgn = sign function: If EFMs are not already binary, reactions were constrained according to the direction (≙ sign) of the flux determined, for zero flux the reaction was considered unconstrained.

36x35mm (300 x 300 DPI)