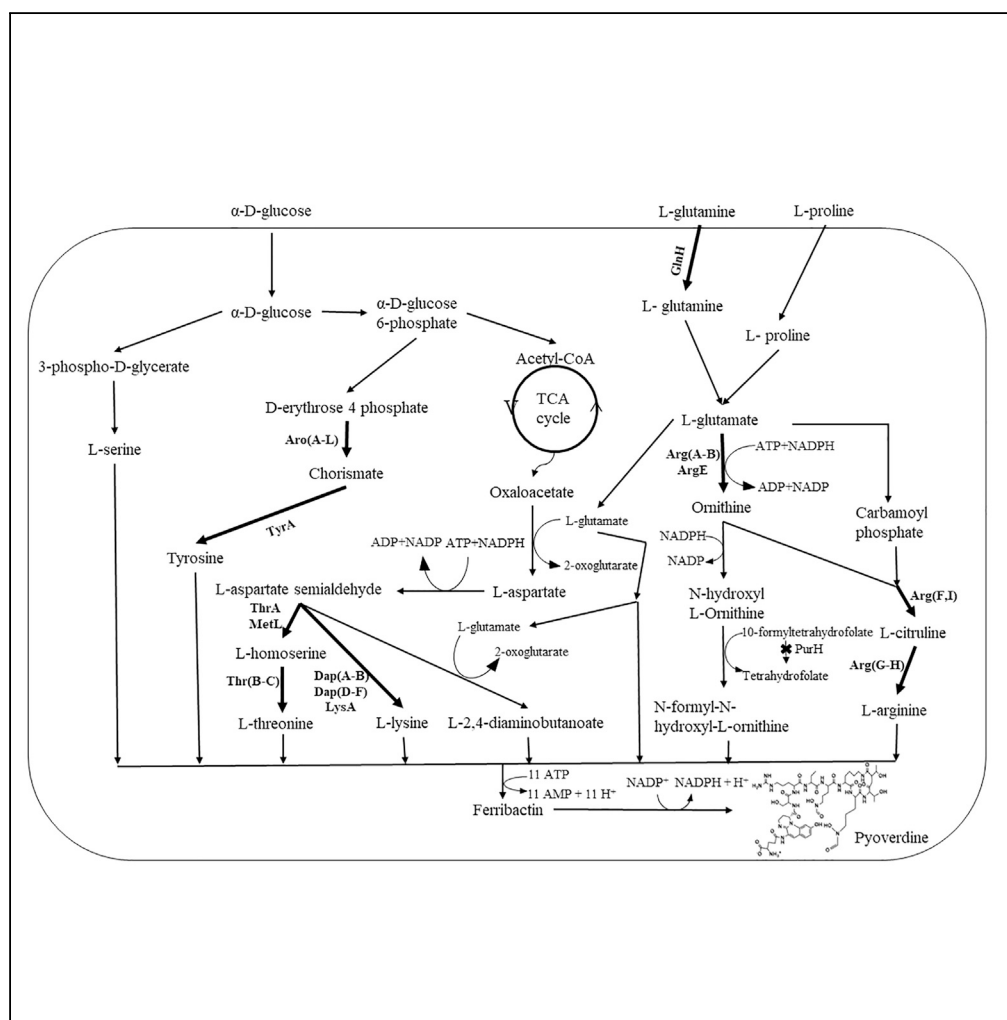## Article

Flux Balance Analysis for Media Optimization and Genetic Targets to Improve Heterologous Siderophore Production

Girish Swayambhu, Nicholas Moscatello, G. Ekin Atilla-Gokcumen, Blaine A. Pfeifer

blainepf@buffalo.edu

HIGHLIGHTS

Metabolic modeling for siderophore heterologous production through *E. coli*

Gene deletion and over-expression targets for five siderophore compounds

Flux balance analysis and Plackett-Burman combination for media optimization

Predicted improvements for latter implementation and siderophore application

Article

Flux Balance Analysis for Media Optimization and Genetic Targets to Improve Heterologous Siderophore Production

Girish Swayambhu,[1] Nicholas Moscatello,[1] G. Ekin Atilla-Gokcumen,[2] and Blaine A. Pfeifer[1,3,*]

SUMMARY

Siderophores are small molecule metal chelators secreted in sparse quantities by their native microbial hosts but can be engineered for enhanced production from heterologous hosts like *Escherichia coli*. These molecules have been proved to be capable of binding heavy metals of commercial and/ or environmental interest. In this work, we incorporated, as needed, the appropriate pathways required to produce several siderophores (anguibactin, vibriobactin, bacillibactin, pyoverdine, and enterobactin) into the base *E. coli* K-12 MG1655 metabolic network model to computationally predict, via flux balance analysis methodologies, gene knockout targets, gene over-expression targets, and media modifications capable of improving siderophore reaction flux. *E. coli* metabolism proved supportive for the underlying production mechanisms of various siderophores. Within such a framework, the gene deletion and over-expression targets identified, coupled with complementary insights from medium optimization predictions, portend experimental implementation to both enable and improve heterologous siderophore production. Successful production of siderophores would then spur novel metal-binding applications.

INTRODUCTION

Siderophores are compounds produced by microbial cells under metal-deficient conditions to sequester metals essential for survival (Miethke and Marahiel, 2007). The metabolic need for siderophores is compounded by the low solubility of metal ions like Fe^{3+} ions and their scarce availability due to innate anatomical binding mechanisms (i.e., heme chelation). Iron accumulation in common minerals as oxides and hydroxides further complicates ready utilization by microorganisms (Kraemer, 2004). Siderophores are often used as a virulence factor by bacteria to outcompete the host's "nutritional immunity" for iron acquisition. For example, transferrin and ferroportin are used to energize iron transport in mammals (Anderson and Vulpe, 2009), and only those bacteria capable of biosynthesizing siderophores that outcompete these mechanisms can establish residence. As such, the strong binding constants of siderophores to heavy metals like iron have been the result of evolutionary pressure.

Given the binding capacity for metal species, there exist several alternative applications for siderophores. Metals in wastewater released into the environment from electroplating, metal finishing, and other heavy manufacturing processes can build to toxic levels within many life forms. A common removal technique is to increase the alkalinity to induce metal precipitation (Kurniawan et al., 2006). This method fails to completely remove trace amounts of metals. Siderophores, due to extremely strong binding abilities (Raymond and Carrano, 1979), can potentially clear water of such toxic metals. Siderophores can also be used as conjugates for stealth antibiotic transport across target microbial pathogen membranes (Ghosh et al., 2017; Schalk, 2018; de Carvalho and Fernandes, 2014), thus potentially circumventing natural bacterial ability to resist antibiotic entry through cell membrane barriers. Apart from this, siderophores may prompt protective immune responses in plants (Aznar et al., 2014) because exposure creates a metal-deficient environment similar to a bacterial attack.

With such diverse and useful applications, siderophores have tremendous impact potential if produced in significant quantities. However, they are only made in trace amounts from their respective native cellular host systems. As a result, heterologous biosynthesis of such metabolites in surrogate cellular systems like *Escherichia coli* and yeast have gained attention (Pfeifer et al., 2003; Zhang et al., 2011). Within these alternative hosts, there are different, and generally more amenable, ways to improve cellular-based metabolite production that include (1) random and targeted mutagenesis and screening for higher production

[1]Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA

[2]Department of Chemistry, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA

[3]Lead Contact

*Correspondence: blainepf@buffalo.edu

https://doi.org/10.1016/j.isci. 2020.101016

mutants, (2) media optimization, and (3) metabolic engineering to address intracellular biosynthetic bottle-necks (Demain, 2006). The first method when conducted at random is experimentally tedious and resource intensive. The remaining methods all benefit from the extended knowledge and advanced engineering tools available to establish heterologous hosts systems, including a range of computational models to better guide product improvement strategies (Orth et al., 2011; Alper et al., 2005).

Many approaches have been developed to model microbial metabolism (King et al., 2015; De Jong et al., 2017). Metabolic networks developed from these models are based on a pseudo-steady-state assumption for ease in calculations (Maarleveld et al., 2013). Such networks are typically underdetermined causing a need for an optimization strategy that is generally linear in nature (Orth et al., 2010). The objective function often chosen to enable a solution for the undetermined network is the flux of the biomass reaction, although this is not the primary target reaction needed to improve product yield (Feist and Palsson, 2010). Flux balance analysis (FBA) is a genome-scale metabolic engineering tool that first establishes and then allows a constrained variation of the carbon flow in a microorganism model based upon steady-state flux distributions with linear flux variations. One issue with FBA is that experimental growth rates resulting from FBA-based perturbations often do not match predictions during early rounds of culturing. A computational solution to this issue was provided through the Minimization of Metabolic Adjustments (MoMA) algorithm in conjunction with FBA (Segre et al., 2002). MoMA additionally accounts for minimum perturbations in flux distributions from the wild-type strain (as a result of genetic changes, for example) and hence reflects the immediate perturbation in output. FBA has also been used in conjunction with other metabolic engineering algorithms to improve product yields. For example, OptForce used with FBA has shown an increased carbon flux toward malonyl-CoA and increased flavanone production by 560% (Xu et al., 2011). Media optimization using the connection between internal metabolic fluxes and external exchange of media components is also an added feature of FBA (Bonarius et al., 1996). Algorithms like OptKnock, Bag of Features, and Genetic Design through Local Search, along with the above-mentioned OptForce, have all been developed to increase the number and precision of gene targets for biosynthetic yield improvement (Burgard et al., 2003; Ranganathan et al., 2010; Lun et al., 2009).

In this work, FBA and its variations have been applied to five siderophore natural products for both media optimization and gene target identification. The structure of the siderophores and their precursors are presented in Figure 1. The results from modeling present the first steps toward a more efficient production system for these siderophores. Subsequent engineering efforts would then be designed to heterologously produce the siderophore compounds and utilize their metal-binding capabilities for a variety of applications.

RESULTS AND DISCUSSION
Single Gene Deletions

We utilized MoMA to identify gene deletion targets for each siderophore. Our objective function contained fluxes of both biomass and siderophore production. To have significant titers (based in part on commensurate biomass accumulation), we restricted our gene deletion targets to only those that have a steady-state growth rate >0.15/h. Our new product fluxes are selected at the exponential growth phase for comparison with the wild-type strain. We categorize the gene targets identified as non-essential or essential genes (or, alternatively, deletions that would lead to auxotrophy), and the top 25 gene deletions for each siderophore were analyzed for increased product formation (Figure 2), although no unique gene deletions were identified for enterobactin. To aid in the visualization of key gene deletion (and over-expression) targets, heterologous siderophore biosynthesis within the context of *E. coli* metabolism is illustrated in Figures 3 and 4.

For anguibactin production, deletion of the *leu* and *glt* genes resulted in the largest increase in product fluxes. This increase in production due to deletion of *leuA* and *gltA* can be explained by consumption of acetyl-CoA (Figure 2; 3-methyl-2-oxobutanoate + acetyl-CoA + H_2O → (2S)-2-isopropylmalate + coenzyme A + H^+; oxaloacetate + acetyl-CoA + H_2O → citrate + coenzyme A + H^+). MoMA also reflected some essential genes like *suc* and *glyA* as potential gene deletions for improved titers. These genes are responsible for conservation of CoA (2-oxo-glutarate + CoA + NAD^+ → succinyl-CoA + CO_2 +NADH) and L-serine (L-serine + tetrahydrofolate → glycine +5,10-methylenetetrahydrofolate + H_2O) for cysteine production, respectively. L-lysine formation is known to consume glucose-6-phosphate in an alternative pathway to fructoselysine phosphate (D-glucose-6-phosphate + L-Lysine → Fructoselysine
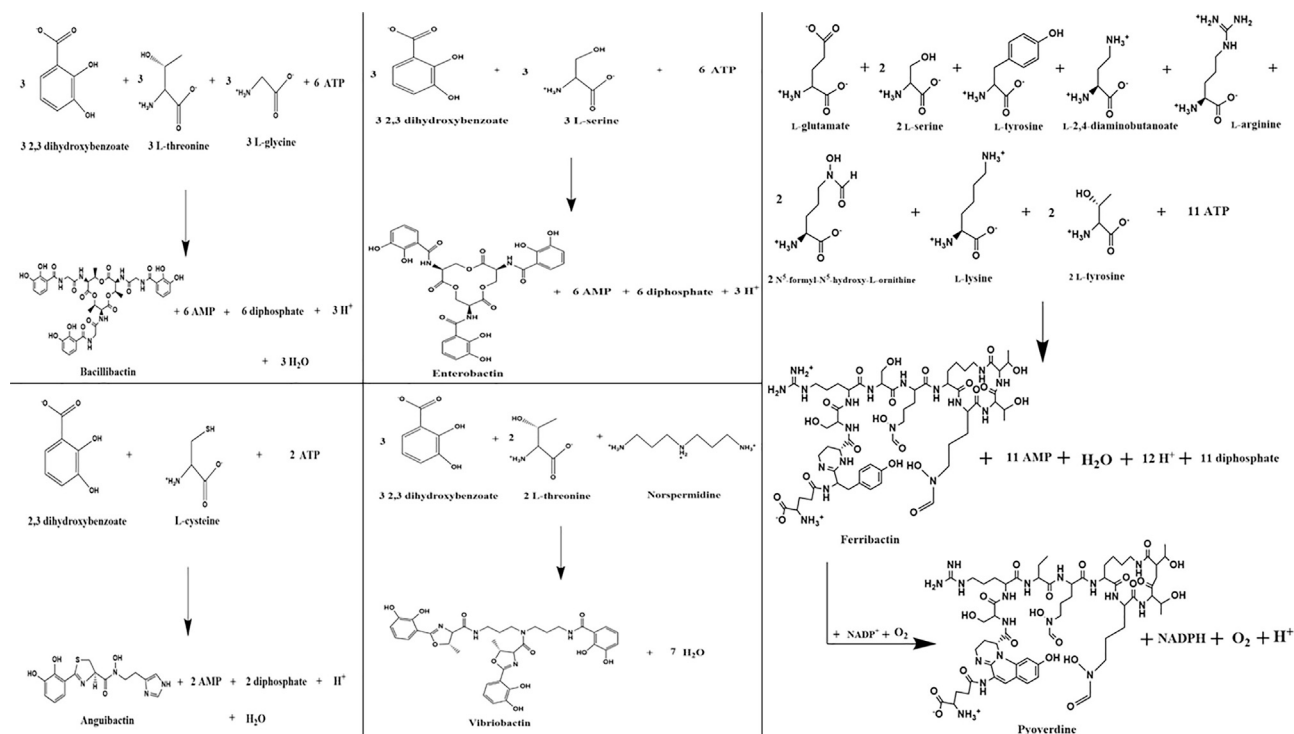
Figure 1. Structure of Siderophores and Their Precursors

6-phosphate + H_2O). As glucose-6-phosphate is essential for L-serine production, deletion of *lysA* and *lysP* improves anguibactin titer. Other acetyl-CoA flux improvement gene deletions include removal of *icd* (more D-threo-isocitrate in the tricarboxylic acid [TCA] cycle for improved acetyl-CoA) and *brnQ* (increased leucine and isoleucine transport into the cytosol to avoid acetyl-CoA consumption). For every molecule of anguibactin, three molecules of ATP are required. Therefore, ATP conservation through deletion of energy-dependent transporter genes *glnH*, *glnP*, and *glnQ*, which encode for L-glutamate transport from the periplasm to cytosol, improves anguibactin titers.

The *leu* gene deletion showed the highest improvements in bacillibactin flux again due to conservation of acetyl-CoA. As three ATP molecules are needed to form a molecule of bacillibactin, ATP conservation is effected through deletion of the *aas* gene (ATP + acyl-carrier protein + 2,3,4-saturated fatty acid → 2,3,4-saturated fatty acyl-[acp] + AMP + diphosphate). MoMA additionally predicted *ptsI* (an essential gene) and *ptsH* gene deletions toward conservation of phosphoenolpyruvate that can make more oxaloacetate essential for L-threonine biosynthesis (phosphoenolpyruvate + N-acetyl-D-glucosamine → N-acetyl-D-glucosamine-6-phosphate + pyruvate). In our models, the availability of glycerol promotes pyruvate formation needed for cell growth; hence deletion of the *ptsI* gene does not affect growth rates. One of the major side reactions that affect bacillibactin production is the consumption of acetyl-CoA for arginine formation, which could otherwise be used for L-threonine biosynthesis. Therefore, prediction of the *argA* gene deletion for enhanced bacillibactin production comes as no surprise (L-glutamate + acetyl-CoA → N-acetyl-L-glutamate + CoA + H^+). The *argA* deletion additionally conserves L-glutamate necessary for L-threonine biosynthesis. The *ilvE* gene encodes for an important rate-limiting step in L-isoleucine biosynthesis that conserves L-glutamate for use in L-threonine biosynthesis ((S)-3-methyl-2-oxopentanoate + L-glutamate → 2-oxoglutarate + L-isoleucine).

Vibriobactin's flux, like the previous two compounds, improves upon deletion of *leu* genes due to conservation of acetyl-CoA useful for L-threonine biosynthesis. The *ilvD* gene deletion reduces 3-methyl-2-oxobutanoate available for reaction with L-glutamate, hence improving L-glutamate flux useful for L-threonine biosynthesis ((R)-2,3-dihydroxy-3-methybutanoate → 3-methyl-2-oxobutanoate + H_2O; 3-methyl-2-oxobutanoate + L-glutamate → L-valine + 2-oxoglutarate). As three ATP molecules are required toward vibriobactin production, deletion of the *ppk*, *adk*, *prs*, and *cmk* genes improves vibriobactin flux. The
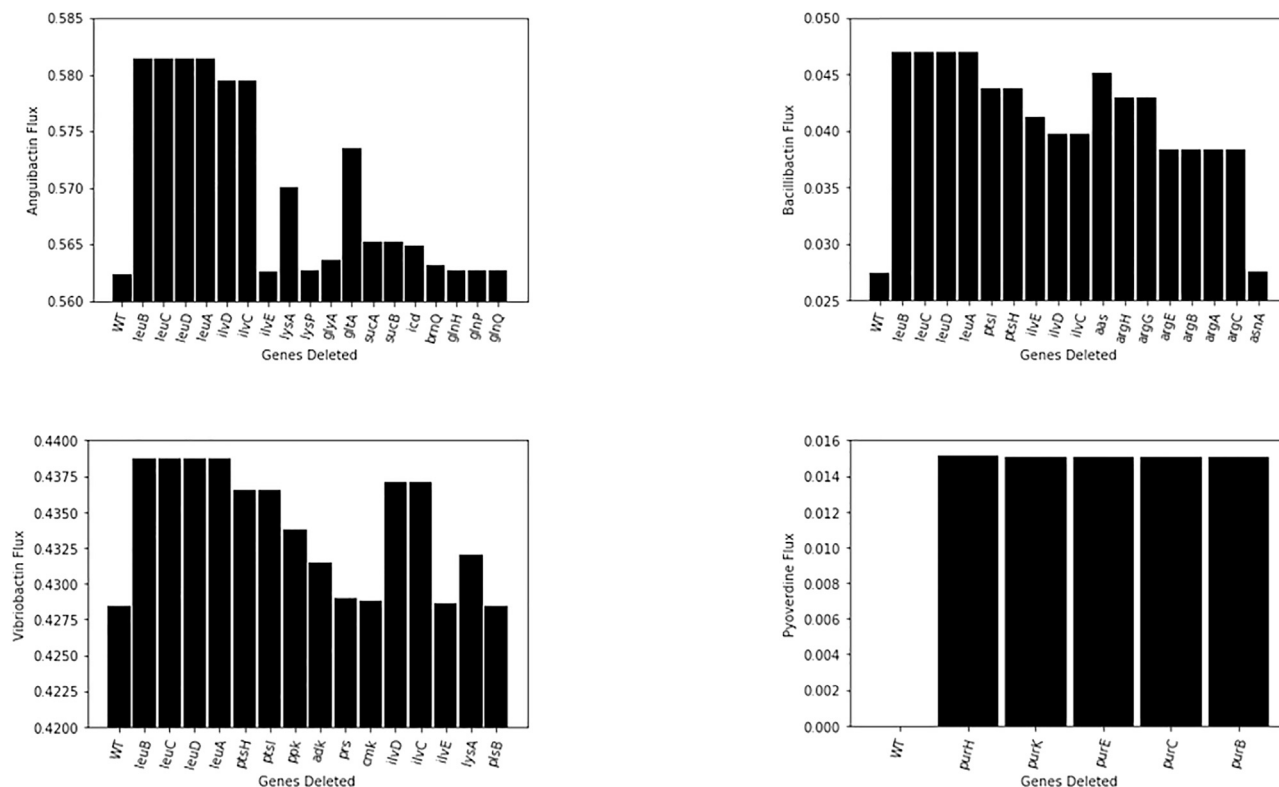
Figure 2. Gene Deletion Targets for Anguibactin, Bacillibactin, Pyoverdine, and Vibriobactin
No gene deletion targets were identified for enterobactin. Flux values are in mmol/gDCW/h and represent compound flux after gene deletion.

ppk and adk genes encode for reactions ATP + diphosphate → ADP + inorganic triphosphate and AMP + ATP → 2 ADP, respectively.

The virulence of *P. aeruginosa* is mainly activated by phosphate starvation, which triggers three different responses, primarily resulting in pyoverdine generation (Zaborin et al., 2009). Thus, in this study we analyze pyoverdine gene deletion targets under phosphate-limiting conditions. There were only five gene deletion targets (*pur* genes) identified. *purH* conserves 10-formyltetrahydrofolate necessary for N-formyl-N-hydroxyl-L-ornithine production (10-formyltetrahydrofolate + 5-amino-1-(5-phospho-D-ribosyl)imidazole-4-carboxamide → tetrahydrofolate + 5-formamido-1-(5-phospho-D-ribosyl)-imidazole-4-carboxamide). Hydrogen carbonate essential for L-arginine biosynthesis is conserved through the deletion of *purK* (5-amino-1-(5-phospho-β-D-ribosyl)imidazole + ATP + hydrogencarbonate → N^5-carboxyaminoimidazole ribonucleotide + ADP + phosphate +2 H^+).

No unique gene deletions were identified for enterobactin. These can be attributed to the depleted environmental or media conditions needed for production in *E. coli*. Enterobactin is known to have a high Fe(III)-binding constant (Carrano and Raymond, 1979) (10^{52} M^{-1}), and hence its production is affected only under extremely low iron concentration conditions (<10^{-52} M).

Different nutrient uptake rates were tested to see if there were any subsequent changes in gene deletions. Glucose and glycerol uptake rates were increased to 10 mmol/gDCW/h and oxygen was increased to 20 mmol/gDCW/h, whereas amino acids were maintained at 0.1 mmol/gDCW/h. The results for anguibactin changed minimally as the *leu* genes still remained the targets that yielded highest anguibactin. However, the *gltA* gene was pushed down in priority and *argA,* which leads to consumed acetyl-CoA toward L-arginine formation, gained predominance in improved anguibactin titers. Results for bacillibactin showed a gene deletion target shift toward *icd* encoding for D-threo-isocitrate + $NADP^+$ → 2-oxo-glutarate + CO_2 + NADPH, which increases acetyl-CoA flux through the TCA cycle. The genes *pheA* and *tyrA* made an entry to the top 25 gene deletion targets owing to an increase of chorismate availability
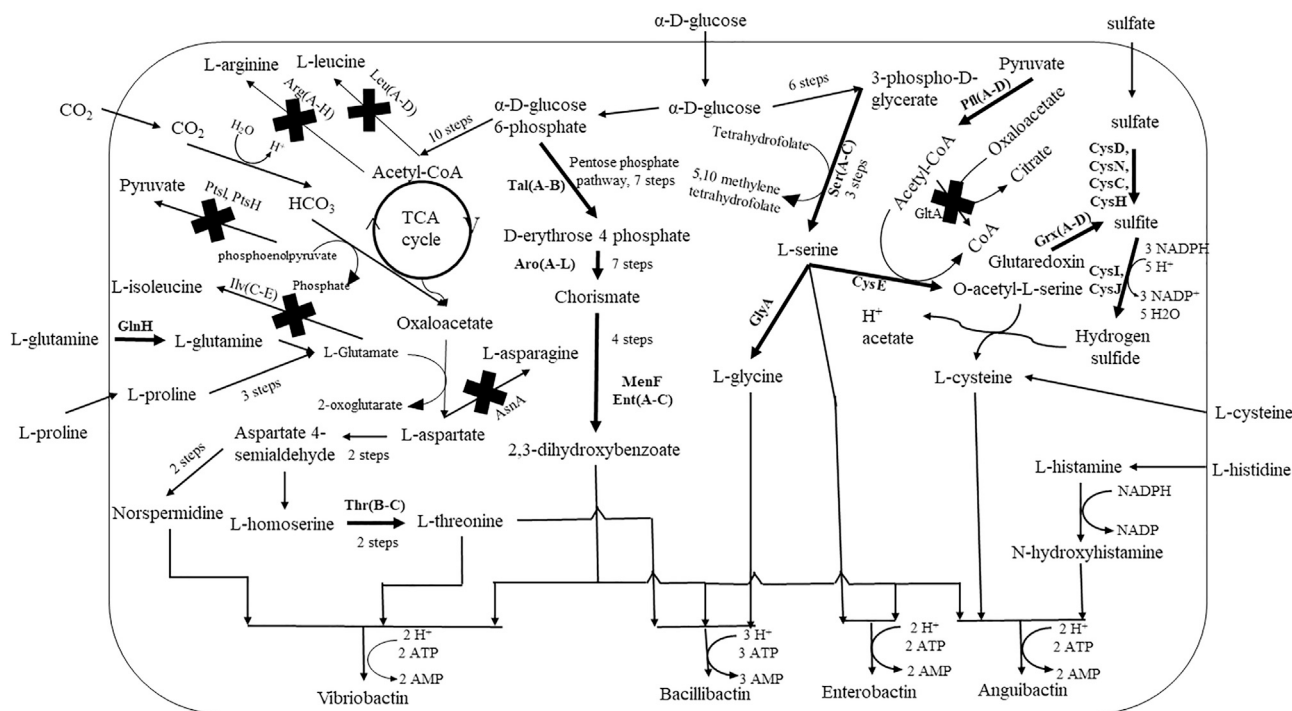
Figure 3. Anguibactin, Bacillibactin, Vibriobactin, and Enterobactin Biosyntheses
Selected gene deletions and over-expressions have been indicated with an **X** and a thickened arrow, respectively.
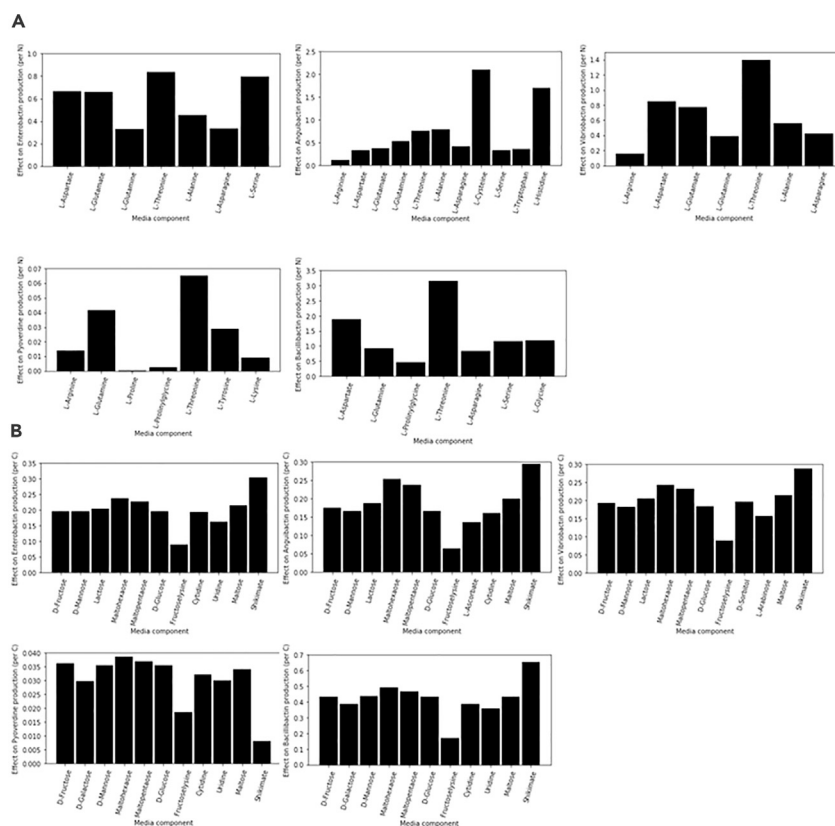
for 2,3-dihydroxybenzoate synthesis (chorismate → prephenate). Vibriobactin showed no major change in the top gene deletion targets. Two additional gene targets were added to the pyoverdine list with nutrient uptake change. The gene *purA*, which encodes for L-aspartate + IMP + GTP → adenylo-succinate + GDP + phosphate + 2H$^+$, conserves L-aspartate necessary for L-threonine formation. Eleven ATP molecules are needed to make one pyoverdine molecule. Hence, the *prs* gene deletion toward ATP conservation is warranted (ATP + D-ribulose-5-phosphate → 5-phospho-α-D-ribose-1-diphosphate + AMP + H$^+$). Overall, increased nutrient uptake increased the link between siderophore production and the TCA cycle, thus causing a change in gene deletion targets.

Gene Over-expressions Using OptForce

OptForce was used to identify gene targets that improved compound production upon over-expression. The algorithm makes use of adjustments of reaction bounds that in turn generate a set of reactions for which the upper limits have to be increased (MUST^U set). The gene over-expression targets that influence a direct precursor are considered as "near pathway," and those that influence the pathway through a series of side reactions are termed as "far pathway." The results are presented in Table S1 with key targets highlighted qualitatively across Figures 3 and 4.

Anguibactin had many "near pathway" over-expression targets. The genes *serA/serB/serC* that encode for L-serine synthesis required for L-cysteine production were identified as one of the primary targets. The *cysC/cysN/cysD/cysI/cysJ* targets that encode for the L-cysteine biosynthesis pathway increase precursor flux. The *purA/purB* and *prs* results are part of the superpathway of histidine biosynthesis, which increases the L-histidine precursor and ATP. ATP has been mentioned earlier to be vital for production of anguibactin. Increased sulfite production for L-cysteine biosynthesis is achieved through the *grxA/grxB/grxC/grxD* targets. In the pentose phosphate pathway, D-erythrose-4-phosphate flux is improved using *talA/talB*. Increased chorismate and, in turn, increased 2,3-dihydroxybenzoate is affected through over-expressions of the *aro* and *ent* genes. Acetyl-CoA increases with an increased flux of 3-phospho-D-glycerate on the glycolysis pathway achieved through over-expression of the *pgk* gene. Some "far pathway" influences include the *ompN/ompC/ompF* genes that encode for various transport mechanisms essential for cell functioning. The genes *nupC/nupG* are responsible for nucleoside and proton transport, which influences

Figure 4. Pyoverdine Biosynthesis
Selected gene deletions and over-expressions have been indicated with an X and a thickened arrow, respectively.

anguibactin as can be seen through media optimizations with identification of cytidine and uridine as flux-improving components.

Bacillibactin has three precursors, L-glycine, 2,3-dihydroxybenzoate, and L-threonine, which help identify "near pathway" targets. The genes *purU/folD* improved tetrahydrofolate production that, in turn, improves L-glycine production. Over-expression of the *aro* genes from the chorismate pathway and *talA/talB* from the pentose phosphate pathway improves 2,3-dihydroxybenzoate production. The target genes *alaA/alaC* improve L-threonine precursor flux. Improved glucose uptake by over-expression of *galP* is identified as a "far pathway" target.

The "near pathway" targets for vibriobactin, apart from the *aro* and *ent* genes, are the *thrA/thrB/thrC* genes, which encode for L-threonine biosynthesis. Glycolysis flux improvement genes like *glk* and *gpmA/gpmM*, which encode for improved glucose-6-phosphate and 3-phospho-D-glycerate, respectively, were identified for improved 2,3-dihydroxybenzoate production. The *tktA/tktB* genes are connected to the pentose phosphate pathway, which upon over-expression would increase D-erythrose-4-phosphate to be utilized through the shikimate pathway to make more 2,3-dihdyroxybenzoate. The "far pathway" targets identified were similar to anguibactin.

Pyoverdine production depends on serine, tyrosine, threonine, lysine, and arginine as important precursors. "Near pathway" targets included *pgl, eno,* and *gapA* genes that improve glycolysis pathway flux for serine and tyrosine generation. L-threonine biosynthesis was improved through *thrA/thrB/thrC* over-expression, whereas L-lysine production was improved via *dapA/dapB/dapD/dapE/dapF* and *lysA* gene targets. The *arg* genes identified as a key gene deletion target with anguibactin and bacillibactin have been identified as one of the prime over-expression targets to improve L-arginine productions for increased pyoverdine flux. In addition to *omp* and *nup*, the *pntA/pntB* genes encoding for hydrogen transport have been identified as "far pathway" targets.

Figure 5. Siderophore Computational Medium Optimization

Media screening results from Plackett-Burman analysis in FBA for (A) amino acids and (B) carbon sources normalized per nitrogen (N) and carbon (C) content.

No unique "near pathway" over-expression targets were identified for enterobactin except the *ent, ser, and aro* genes that influence 2,3-dihydroxybenzoate and L-cysteine fluxes. The *omp* genes were identified as "far pathway" gene targets.

Media Optimization

There are many optimization strategies that reduce the number of experimental designs required to enhance media compositions (Kennedy and Krouse, 1999). Without these strategies, a simple trial-and-error approach would be time consuming and expensive. FBA allows us to constrain the media components within a stoichiometric model. Thus, FBA can be used to computationally predict an enhanced medium composed of amino acids, carbon sources, vitamins, and other nutrients that can then be verified experimentally.

In this work, we have used a Plackett-Burman-based screening comprising the analysis of 7 or 11 components at a time. The screening was applied across nearly 100 components (mole basis) on all the siderophores in this study, and the best media components were hand-picked using statistical criteria (outlined in the Methods section). Tables S2 and S3 demonstrate the general outline the screening utilized for these calculations. Figure 5 shows the various carbon sources and amino acids that most contributed toward improved production.

The carbon source that influenced maximum siderophore production was shikimate, an important part of the chorismate pathway essential in many siderophore biosynthetic pathways. This carbon source improves 2,3-dihydroxybenzoate flux, which is one of the precursors needed to make anguibactin, vibriobactin, enterobactin, and bacillibactin. As 2,3-dihydroxybenzoate is not essential for production of pyoverdine, shikimate does not influence pyoverdine production levels. Among more conventional carbon sources, maltohexaose and maltopentaose showed maximum influence on all siderophore products when

compared with D-glucose or D-fructose. Apart from these, lactose and maltose showed increased production of all siderophores. Maltohexaose and maltopentaose are expensive compounds (relative to standard medium carbon sources), and hence lactose and maltose may serve as more economical options for increased production.

Anguibactin production showed more pronounced improvements upon increased L-cysteine and L-histidine and minor improvements upon increased L-threonine and L-alanine among amino acids. L-cysteine in one of the direct precursors essential for anguibactin production. L-histidine is known to improve N-hydroxyhistamine, which is one of the precursors for anguibactin biosynthesis (Figure 3). L-threonine is degraded to 2-oxobutanoate, which further degrades to succinyl-CoA to improve acetyl-CoA through the TCA cycle. L-alanine can release CoA from pimeloyl-CoA (S-pimeloyl derivative of coenzyme A), which improves acetyl-CoA flux.

Major amino acids to improve bacillibactin production include L-threonine, L-aspartate, L-glycine, and L-serine. L-threonine is a precursor for bacillibactin production. L-glycine is a precursor for bacillibactin, and hence L-serine, which is converted to L-glycine, serves as an important media component to boost production. The pathway conversion of oxaloacetate to L-threonine includes amino acids like L-aspartate (Figure 3).

The media components identified for vibriobactin are similar to bacillibactin due to two matching precursors, L-threonine and 2,3-dihydroxybenzoate. In addition, norspermidine made from L-aspartate is an important precursor that explains the identification of L-aspartate and L-glutamate as parameters to improve vibriobactin production.

Enterobactin is made from L-serine and 2,3-dihydroxybenzoate. There is no TCA cycle/acetyl-CoA involvement in its biosynthesis. Therefore the identification of L-threonine, L-aspartate, and L-glutamate as boosting media components is unexpected. One possible explanation could be degradation of L-threonine (which can be made from L-aspartate and L-glutamate) into L-glycine, which in turn converts to L-serine.

The identification of L-threonine, L-tyrosine, L-lysine, and L-arginine as production boosters for pyoverdine comes as no surprise, as all these amino acids are biosynthetic precursors. L-glutamine has also been identified due to its importance in the formation of L-2,4-diaminobutanoate and N-formyl-N-hydroxyl-L-ornithine (Figure 4). Additional source code and raw data for the computational approaches of this study are included as Data S1.

Limitations of the Study

Steady-state FBA has disadvantages when compared with real-time variable flux values within microorganisms. This includes, to some extent, the actual effects on growth rates, which must be verified experimentally upon the implementation of predicted targets and, theoretically, should be mitigated by using approaches such as MoMA. The steady-state approach attempts to mimic an average growth flux throughout the organism's life cycle to predict the net product yield. In reality, a living microorganism continues to have flux changes throughout its life cycle, and hence a dynamic steady-state FBA is more accurate. However, the constraints of memory and computational time make it difficult to solve the multiple differential equations required for dynamic FBA.

In a similar fashion, stoichiometric models cannot provide insight into metabolism kinetics or associated regulatory networks. Regulation in particular is strongly associated with secondary metabolite natural products (with external signals shifting cellular biosynthesis). As siderophores share biosynthetic steps associated with secondary metabolite natural products, they will likely be subjected to the same regulatory elements dictating eventual biosynthesis. This issue is lessened when biosynthesis is established within a distant heterologous host, as we describe in this study, because the new host is often chosen to minimize overlapping metabolic and regulatory networks with the native host in an attempt to provide a "cleaner" slate with which engineering can be applied, and as a result, stoichiometric modeling may be more accurate. Likewise, upon implementation of model-based predictions within a heterologous host, regulation elements can be minimized further via user-directed constitutive or inducible gene expression elements specific to the new host.

Conclusions

FBA offers a computational advantage of identifying gene targets and ideal media components by effectively coupling with various algorithms that could otherwise be difficult to analyze experimentally. In short, the approach offers an alternative to a purely experimental route to boost production of natural products. In this study, we identified multiple computational gene targets (deletions and over-expressions) and media parameters to improve production of a variety of siderophore natural products. The concepts of stoichiometric analysis of reactions and genes that are associated with siderophore biosynthesis have been effectively executed with linear programming to economically predict best cellular performance. Future steps would be to experimentally implement the identified gene targets and media component changes for improved heterologous siderophore production in anticipation of a variety of novel applications.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101016.

AUTHOR CONTRIBUTIONS

G.S. and B.A.P. conceived the study and prepared the manuscript. G.S. conducted the research while N.M. and G.E.A.-G. provided technical support.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Alper, H., Jin, Y.-S., Moxley, J., and Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. Metab. Eng. *7*, 155–164.

Anderson, G.J., and Vulpe, C.D. (2009). Mammalian iron transport. Cell. Mol. Life Sci. *66*, 3241.

Aznar, A., Chen, N.W., Rigault, M., Riache, N., Joseph, D., Desmaële, D., Mouille, G., Boutet, S., Soubigou-Taconnat, L., and Renou, J.-P. (2014). Scavenging iron: a novel mechanism of plant immunity activation by microbial siderophores. Plant Physiol. *164*, 2167–2183.

Bonarius, H.P., Hatzimanikatis, V., Meesters, K.P., de Gooijer, C.D., Schmid, G., and Tramper, J. (1996). Metabolic flux analysis of hybridoma cells in different culture media using mass balances. Biotechnol. Bioeng. *50*, 299–318.

Burgard, A.P., Pharkya, P., and Maranas, C.D. (2003). Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. *84*, 647–657.

Carrano, C.J., and Raymond, K.N. (1979). Ferric ion sequestering agents. 2. Kinetics and mechanism of iron removal from transferrin by enterobactin and synthetic tricatechols. J. Am. Chem. Soc. *101*, 5401–5404.

de Carvalho, C.C., and Fernandes, P. (2014). Siderophores as "Trojan Horses": tackling multidrug resistance? Front. Microbiol. *5*, 290.

De Jong, H., Casagranda, S., Giordano, N., Cinquemani, E., Ropers, D., Geiselmann, J., and Gouzé, J.-L. (2017). Mathematical modelling of microbes: metabolism, gene expression and growth. J. R. Soc. Interface *14*, 20170502.

Demain, A.L. (2006). From natural products discovery to commercialization: a success story. J. Ind. Microbiol. Biotechnol. *33*, 486–495.

Feist, A.M., and Palsson, B.O. (2010). The biomass objective function. Curr. Opin. Microbiol. *13*, 344–349.

Ghosh, M., Miller, P.A., Möllmann, U., Claypool, W.D., Schroeder, V.A., Wolter, W.R., Suckow, M., Yu, H., Li, S., and Huang, W. (2017). Targeted antibiotic delivery: selective siderophore conjugation with daptomycin confers potent activity against multidrug resistant *Acinetobacter baumannii* both in vitro and in vivo. J. Med. Chem. *60*, 4577–4583.

Kennedy, M., and Krouse, D. (1999). Strategies for improving fermentation medium performance: a review. J. Ind. Microbiol. Biotechnol. *23*, 456–475.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., Ebrahim, A., Palsson, B.O., and Lewis, N.E. (2015). BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. *44*, D515–D522.

Kraemer, S.M. (2004). Iron oxide dissolution and solubility in the presence of siderophores. Aquat. Sci. *66*, 3–18.

Kurniawan, T.A., Chan, G.Y.S., LO, W.H., and Babel, S. (2006). Physico-chemical treatment techniques for wastewater laden with heavy metals. Chem. Eng. J. *118*, 83–98.

Lun, D.S., Rockwell, G., Guido, N.J., Baym, M., Kelner, J.A., Berger, B., Galagan, J.E., and Church, G.M. (2009). Large-scale identification of

genetic design strategies using local search. Mol. Syst. Biol. 5, 296.

Maarleveld, T.R., Khandelwal, R.A., Olivier, B.G., Teusink, B., and Bruggeman, F.J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. Biotechnol. J. 8, 997–1008.

Miethke, M., and Marahiel, M.A. (2007). Siderophore-based iron acquisition and pathogen control. Microbiol. Mol. Biol. Rev. 71, 413–451.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.Ø. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. Mol. Syst. Biol. 7, 535.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. 28, 245.

Pfeifer, B.A., Wang, C.C., Walsh, C.T., and Khosla, C. (2003). Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using Escherichia coli as a heterologous host. Appl. Environ. Microbiol. 69, 6698–6702.

Ranganathan, S., Suthers, P.F., and Maranas, C.D. (2010). OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6, e1000744.

Raymond, K.N., and Carrano, C.J. (1979). Coordination chemistry and microbial iron transport. Acc. Chem. Res. 12, 183–190.

Schalk, I.J. (2018). Siderophore-antibiotic conjugates: exploiting iron uptake to deliver drugs into bacteria. Clin. Microbiol. Infect. 24, 801.

Segre, D., Vitkup, D., and Church, G.M. (2002). Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U S A 99, 15112–15117.

Xu, P., Ranganathan, S., Fowler, Z.L., Maranas, C.D., and Koffas, M.A. (2011). Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab. Eng. 13, 578–587.

Zaborin, A., Romanowski, K., Gerdes, S., Holbrook, C., Lepine, F., Long, J., Poroyko, V., Diggle, S.P., Wilke, A., and Righetti, K. (2009). Red death in Caenorhabditis elegans caused by Pseudomonas aeruginosa PAO1. Proc. Natl. Acad. Sci. U S A 106, 6327–6332.

Zhang, H., Boghigian, B.A., Armando, J., and Pfeifer, B.A. (2011). Methods and options for the heterologous production of complex natural products. Nat. Product Rep. 28, 125–151.

Supplemental Information

Flux Balance Analysis for Media

Optimization and Genetic Targets to Improve

Heterologous Siderophore Production

Girish Swayambhu, Nicholas Moscatello, G. Ekin Atilla-Gokcumen, and Blaine A. Pfeifer

Transparent Methods:

A genome scale stoichiometric model of *E. coli* IJO1366 was downloaded from the BiGG models database (www.<u>bigg.ucsd.edu/models/iJO1366</u>). To incorporate reactions for heterologous production mechanisms for anguibactin, bacillibactin, vibriobactin, and pyoverdine into the IJO1366 model, we utilized the 'AddReaction' function (note: enterobactin is native to *E. coli*). Reactions needed to complete the production cycle for each siderophore were taken from <u>http://biocyc.org/</u>. For anguibactin, the starting molecules like 2,3-dihydroxybenzoate and L-cysteine are already part of the *E. coli* genome while N-hydroxyhistamine was added using two reactions from histidine: 1) L-histidine + H^+ → CO_2 + histamine; 2) histamine + NADPH + oxygen → N-hydroxyhistamine + $NADP^+$ + H_2O. Bacillibactin is formed through reaction of 2,3-dihydroxybenzoate adenylate with adenylated products of L-threonine and glycine: 1) glycine + ATP → (glycyl)adenylate + diphosphate; 2) L-threonine + ATP + H^+ → (L-threonyl)adenylate + diphosphate; 3) 2,3-dihydroxybenzoate adenylate + (glycyl)adenylate + (L-threonyl)adenylate → 4 H^+ + 3 AMP + bacillibactin. For vibriobactin, reactions were input to build norspermidine from L-aspartate 4-semialdehyde as follows: 1) L-aspartate 4-semialdehyde + L-glutamate → L-2,4-diaminobutanoate + 2-oxoglutarate; 2) L-2,4-diaminobutanoate + H^+ → propane-1,3-diamine + CO_2; 3) L-aspartate 4-semialdehyde + propane-1,3-diamine + NADPH + H^+ → carboxyspermidine + $NADP^+$ + H_2O; 4) carboxyspermidine + H^+ → norspermidine + CO_2. One mole of norspermidine combines with two moles of L-threonine and three moles of 2,3-dihydroxybenzoate to form one mole of vibriobactin as follows: norspermidine + 2 L-threonine + 3 2,3-dihydroxybenzoate → vibriobactin + 7 H_2O. Pyoverdine requires eight precursor molecules. Out of these, six are amino acids (L-arginine, L-glutamate, L-lysine, L-serine, L-threonine, and L-tyrosine), and L-2,4-diaminobutanoate is generated via the route introduced above for vibriobactin. N^5-formyl-N^5-hydroxy-L-ornithine is generated using the following reactions: 1) L-ornithine + NADPH + oxygen → N^5-hydroxy-L-ornithine + $NADP^+$ + H_2O; 2) N^5-hydroxy-L-ornithine + 10-formyltetrahydrofolate → N^5-formyl-N^5-hydroxy-L-ornithine + tetrahydrofolate.

Calculations were made in MATLAB R2018a using the 4.1 SBML toolbox and the 2017 COBRA (Constraint Based Reconstruction and Analysis) toolbox. Optimization was conducted using the Gurobi solver v8.1 interfaced with MATLAB and COBRA. Source code and raw data have been included as additional supplementary data.

Flux Balance Analysis:

Flux Balance Analysis (FBA) is a mathematical approach towards predicting the flow of metabolites through an organism's cellular mathematical model. The major use of FBA is to identify optimal points in a constrained solution space based upon the sets of metabolites that need to be optimized. For example, lipids from algae can be predicted for optimum value based on setting an objective function as a set of biomass and lipid flux values with different coefficients for each of their producing reactions. Metabolic reactions are represented as a matrix (S) of size m*n. Every row represents a compound and every column represents a reaction. Since the overall mass balance under a steady state assumption is zero, we have

$$S \cdot \boldsymbol{v} = 0 \qquad\qquad (1)$$

where $\boldsymbol{v}$ is a matrix representing the flux through each reaction subject to a constraint,

$$a_i < \boldsymbol{v}_i < b_i \qquad\qquad (2)$$

where a_i and b_i represent lower and upper reaction flux bounds, respectively.

Although an extensive solution space is available, FBA is able to identify a single optimal distribution that lies on the edge of the allowable solution space. This is done through an objective function which gives the FBA program a set of coefficients assigned to each reaction according to their need for optimization. For example, an objective function could purely be a biomass flux optimizing equation with the coefficient of biomass set to one. Otherwise, it could include a series of other coefficients for all metabolite producing reactions that need to be optimized. The objective function is represented as:

$$Z = C^T \boldsymbol{v} \qquad\qquad\qquad (3)$$

where C is a row vector containing the coefficients for each individual reaction in the objective function and T represents the vector transpose. Equation 1 is generally underdetermined because there are more unknown flux values than available reaction equations in large scale models(Varma and Palsson, 1994). This results in a large solution space with no unique solutions. FBA is one method that can identify optimal points within this constraint space. Linear programming is used for such optimizations(Orth et al., 2010). Apart from equation 1, we have lower and upper bounds for reaction flux v and linear combination of fluxes (objective function) that can be used as equations to solve for a set of flux values which is output in FBA. This output either maximizes or minimizes the objective function.

Minimization of Metabolic Adjustment (MoMA):

MoMA was developed to predict the immediate effect of a gene deletion with more accuracy as compared to FBA(Shlomi et al., 2005). FBA cannot showcase the effect of its predicted gene deletion until about 700 generations later. In contrast, MoMA is designed to avoid this by causing minimum perturbations that can improve the optimizing parameter, allowing effects to emerge immediately or in fewer generations. This is formulated as a quadratic programming problem:

Minimize: $(x - w)^T (x - w)$, subject to

$$S \cdot \boldsymbol{v} = 0 \text{ and } a_i < v_i < b_i \qquad\qquad (4)$$

Where w is the wild type flux vector and x is the deletion strain flux vector (other variables introduced above). We can generally analyze one gene deletion at a time or focus on two gene deletions at a time. In this work, we have focused on single gene deletions only to avoid extensive computation times.

OptForce:

OptForce is an algorithm that does not directly involve flux calculations but changes the lower and upper bounds of reaction fluxes to improve product titers(Ranganathan et al., 2010). Those reaction

fluxes that must increase for over-production are marked as MUST^U sets and those fluxes that must decrease for over-production are marked as MUST^L sets. We ignored the MUST^L sets since they are already accounted for in gene deletions. Thus, only the MUST^U set was considered for over-expression targets.

Media Optimization:

A Plackett and Burman design of experiments was conducted with different carbon and nitrogen sources(Plackett and Burman, 1946). While analyzing the effect of different nitrogen sources, we uniformly set a glucose and glycerol uptake rate of 3 mmol/gDCW/hr while oxygen is set at 8 mmol/gDCW/hr according to experimental uptake rates under glucose limited conditions(Weaver et al., 2014). While analyzing the effect of different carbon sources, we uniformly set a basic nitrogen source comprised of 3 mmol/gDCW/hr each of L-threonine, L-cysteine, L-histidine, and L-glutamine along with an oxygen uptake rate of 8 mmol/gDCW/hr. The maximum and minimum uptake rates for different parameters were set at 3 and 0.1 mmol/gDCW/hr, respectively, according to experimentally observed values for most amino acids(Selvarasu et al., 2009).

Tables S1 and S2 represents scenarios of different media components tested seven (Figure 5A x-axis) or 11 (Figure 5B x-axis) at a time, respectively. A '-' condition represents a flux of 0.1 mmol/gDCW/hr and a '+' condition represents a flux of 3 mmol/gDCW/hr. A flux balance analysis was conducted to calculate net product titer for a particular set of media components. The effect of every component in a scenario was calculated as sums of all '+' conditions and differences of all '-' conditions and net flux values divided by total number of components.

$$\text{Effect} = \frac{(\sum_{i=1}^{n} positive\ conditions - \sum_{i=1}^{n} negative\ conditions)}{ntotal * (number\ of\ C\ or\ N\ per\ mole)} \quad (5)$$

Only those components in a scenario that had effects higher than the formula below were considered as promising media optimization targets,

$$Cutoff = Mean + 0.75 * SD$$

Tables

Table S1. Over-expression Targets via OptForce, Related to Figures 3 and 4. Over-expression targets for different siderophores inclusive of "near pathway" targets that influence siderophore or precursor production directly or "far pathway" targets that do not directly influence precursor production.

Near pathway over-expressions

Anguibactin				Vibriobactin				Pyoverdine				Bacillibactin				Enterobactin			
Gene(s)	Rationale 1	Pathway	Rationale 2	Gene(s)	Rationale 1	Pathway	Rationale 2	Gene(s)	Rationale 1	Pathway	Rationale 2	Gene(s)	Rationale 1	Pathway	Rationale 2	Gene(s)	Rationale 1	Pathway	Rationale 2
ser(A-C)	Improved L-Serine production	L-serine biosynthesis	Increased L-cysteine production	glk	Improved glucose 6-phosphate	Glycolysis	Improved 2,3 dihydroxybenzoate biosynthesis	pgl	Improved glucose 3 phosphate	Glycolysis	Improved serine, tyrosine and acetyl coA	alaA, alaC	Improved glutamate production	L-alanine biosynthesis	Improved L-threonine	entA, entD, entE, entF	More 2,3 dihydroxybenzoate	Enterobactin biosynthesis	
prs	Improved 5-phospho-alphaD ribose 1 diphosphate	Super pathway of histidine, purine, and pyramidine biosynthesis	Increased L-histidine production	thr(A-C)	Improved L-threonine	L-Threonine biosynthesis		eno	improved pyruvate	glycolysis	Improved acetyl coA	folD	Improved tetrahydrofolate synthesis	N-formyl tetrahydrofolate biosynthesis	Improved glycine	ser(A-C)	Improved L-Serine production	L-serine biosynthesis	Increased L-cysteine production
grx(A-D)	Increase sulfite production	L-cysteine biosynthesis	Increased L-cysteine production	ackA	Improved Acetyl phosphate	acetate and ATP formation from acetyl coA	Improved acetyl coA production	gapA	Improved 3 phospho D glyceroyl phosphate	Glycolysis	Improved serine, tyrosine and acetyl coA	aro(A-L)	Increase chorismate production	Chorismate pathway	Increased 2,3-dihydroxybenzoate	aro(A-L)	Increase Chorismate production	Chorismate pathway	Increased 2,3-dihydroxybenzoate
cysE	Increased L-cysteine production	L-cysteine biosynthesis		gpmA,gpmM	Improved 3 phospho D glycerate	Glycolysis	Improved acetyl coA & 2,3 dihydroxybenzoate biosynthesis	dap(A-B)	Improved tetrahydrodipicilonate	L-lysine biosynthesis	Improved L-lysine	purU	Improved tetrahydrofolate synthesis	formyltetrahydrofolate deformylase	Improved glycine				
cysC	Increased L-cysteine production	L-cysteine biosynthesis		tktA,tktB	Improved glucose 3 phosphate	Pentose phosphate pathway	Improved acetyl coA & 2,3 dihydroxybenzoate biosynthesis	dap(D-F)	Improved meso diaminopimelate	L-lysine biosynthesis	Improved L-lysine	talA, talB	Improved D-erythrose 4 phosphate	pentose phosphate pathway	Improved 2,3 dihydroxybenzoate synthesis				
cysN,cysD	Increased L-cysteine production	L-cysteine biosynthesis		pgk	Improved 3 phospho D glycerol phosphate biosynthesis	Glycolysis	Improved acetyl coA & 2,3 dihydroxybenzoate biosynthesis	lysA	Improved L-lysine	L-lysine biosynthesis		glyA	Improved glycine production	Glycine biosynthesis					
cysJ,cysI	Increased L-cysteine production	L-cysteine biosynthesis		aro(A-L)	Increase Chorismate production	Chorismate pathway	Increased 2,3-dihydroxybenzoate	thr(A-C)	Improved L-threonine	L-Threonine biosynthesis		ser(A-C)	Improved L-Serine production	L-serine biosynthesis	Increased L-cysteine production				
tal(A-B)	Increased D-erythrose 4 phosphate	Pentose phosphate pathway	Increased 2,3-dihydroxybenzoate	ent(A-C)	Increase 2,3-dihydroxybenzoate	2,3-dihydroxybenzoate biosynthesis		arg(A-B), argE	Improved Ornithine	L-Arginine biosynthesis	Improved N-formyl-N-hydroxy-L-ornithine	gpmA,gpmM,tktA,tktB	Improved 3 phospho D glycerate	Glycolysis/Pentose phosphate pathway	Improved acetyl coA & 2,3 dihydroxybenzoate biosynthesis				
aro(A-L)	Increase Chorismate production	Chorismate pathway	Increased 2,3-dihydroxybenzoate					arg(F-H)	Improved L-Arginine	L-Arginine biosynthesis									
ent(A-C)	Increase 2,3-dihydroxybenzoate	2,3-dihydroxybenzoate biosynthesis						aro(A-L)	Increase Chorismate production	Chorismate pathway	Increased 2,3-dihydroxybenzoate								
pgk	Increased 3 phospho D glycerate	Glycolysis	Increased L-cysteine production																
purA, purB	Improved ADP/ATP production	Super pathway of histidine, purine, and pyramidine biosynthesis	Increased NADPH/N-hydroxyhistamine																
purH	Improved tetrahydrofolate synthesis	Inosine-5'-phosphate biosynthesis	Increased L-cysteine production																

Far pathway over-expressions

Anguibactin		Vibriobactin		Pyoverdine		Bacillibactin		Enterobactin	
Gene(s)	Rationale 1	Gene(s)	Rationale 1	Gene(s)	Rationale 1	Gene(s)	Rationale 1	Gene(s)	Rationale 1
ompN,ompC,ompF nupC,nupG	Outer membrane porins Nucleoside and H+ transporter	ompN,ompC,ompF nupC,nupG	Outer membrane porins Nucleoside and H+ transporter	pntA,pntB nupC,nupG	H+ transporter Nucleoside and H+ transporter	nupC,nupG ompN,ompC,ompF,phoE galP	Nucleoside and H+ transporter Acetate transport Increased glucose intake into cytochrome	ompN,ompC,ompF,phoE	Outer membrane porins H transport
galP	More glucose			ompN,ompC,ompF,phoE	Outer membrane porins H transport				
gnd	More NADPH								

Table S2. Medium Optimization Design (Primarily for Amino Acid Sources), Related to Figure 5. Plackett-Burman screening matrix for analysis of seven media components at a time. C1 through C7 denote Components 1 to 7, and 1-8 in column 1 indicate different media formulations. Uptake rates are in mmol/gDCW/hr.

	C1	C2	C3	C4	C5	C6	C7
1	+	-	-	+	-	+	+
2	+	+	-	-	+	-	+
3	+	+	+	-	-	+	-
4	-	+	+	+	-	-	+
5	+	-	+	+	+	-	-
6	-	+	-	+	+	+	-
7	-	-	+	-	+	+	+
8	-	-	-	-	-	-	-
	+	3					
	-	0.1					

Table S3. Medium Optimization Design (for Carbon Sources), Related to Figure 5. Plackett-Burman screening matrix for analysis of eleven media components at a time. C1 through C11 denote Components 1 to 11, and 1-12 in column 1 indicate different media formulations. Uptake rates are in mmol/gDCW/hr.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
1	+	+	+	+	+	+	+	+	+	+	+
2	+	+	+	+	-	-	-	+	-	-	-
3	+	+	-	-	-	+	-	-	+	-	+
4	+	-	+	-	+	+	+	-	-	-	-
5	+	-	-	+	-	-	+	-	+	+	-
6	+	-	-	-	+	-	-	+	-	+	+
7	-	+	+	-	-	-	+	-	-	+	+
8	-	+	-	+	+	+	-	-	-	+	-
9	-	+	-	-	+	-	+	+	+	-	-
10	-	-	+	+	+	-	-	-	+	-	+
11	-	-	+	-	-	+	-	+	+	+	-
12	-	-	-	+	-	+	+	+	-	-	+
	+	3									
	-	0.1									

Supplementary References

ORTH, J. D., THIELE, I. & PALSSON, B. Ø. 2010. What is flux balance analysis? *Nature Biotechnology,* 28**,** 245.

PLACKETT, R. L. & BURMAN, J. P. 1946. The design of optimum multifactorial experiments. *Biometrika,* 33**,** 305-325.

RANGANATHAN, S., SUTHERS, P. F. & MARANAS, C. D. 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Computational Biology,* 6**,** e1000744.

SELVARASU, S., OW, D. S. W., LEE, S. Y., LEE, M. M., OH, S. K. W., KARIMI, I. A. & LEE, D. Y. 2009. Characterizing *Escherichia coli* DH5α growth and metabolism in a complex medium using genome-scale flux analysis. *Biotechnology and Bioengineering,* 102**,** 923-934.

SHLOMI, T., BERKMAN, O. & RUPPIN, E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences,* 102**,** 7695-7700.

VARMA, A. & PALSSON, B. O. 1994. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.,* 60**,** 3724-3731.

WEAVER, D. S., KESELER, I. M., MACKIE, A., PAULSEN, I. T. & KARP, P. D. 2014. A genome-scale metabolic flux model of *Escherichia coli* K–12 derived from the EcoCyc database. *BMC Systems Biology,* 8**,** 79.