Metabolic modeling of *Streptococcus mutans* reveals complex nutrient requirements of an oral pathogen

Kenan Jijakli and Paul A. Jensen[#]

Department of Bioengineering and Carl R. Woese Institute for Genomic Biology
University of Illinois at Urbana-Champaign, Urbana, IL, USA

Running title: Metabolic modeling of *S. mutans*.

[#]Correspondence to: Paul A. Jensen, pjens@illinois.edu

Abstract

Streptococcus mutans is a Gram positive bacterium that thrives under acidic conditions and is a primary cause of tooth decay (dental caries). To better understand the metabolism of *S. mutans* on a systematic level, we manually constructed a genome-scale metabolic model of the *S. mutans* type strain UA159. The model, called iSMU, contains 656 reactions involving 514 metabolites and the products of 488 genes.

We interrogated *S. mutans*' nutrient requirements using model simulations and nutrient removal experiments in defined media. The iSMU model matched experimental results in greater than 90% of the conditions tested. We also simulated effects of single gene deletions. The model's predictions agreed with 78.1% and 84.4% of the gene essentiality predictions from two experimental datasets. Our manually curated model is more accurate than *S. mutans* models generated from automated reconstruction pipelines. We believe the iSMU model is an important resource for understanding how metabolism enables the cariogenicity of *S. mutans*.

Introduction

Streptococcus mutans is one of over 600 species of bacteria in the oral microbiome. This Gram-positive, lactic acid bacterium thrives in the oral environment in part due to its metabolic flexibility. *S. mutans* can feed on several carbohydrates (1) and has complex, interdependent amino acid auxotrophies (2). *S. mutans* is the primary cause of tooth decay (dental caries). By fermenting a wide array of dietary sugars into lactic acid, *S. mutans* creates a highly acidic microenvironment near the tooth surface (as low as pH 3.0). The lactic acid demineralizes the tooth structure, resulting in decay.

Understanding the acidogenic capabilities of the *S. mutans* requires an unbiased, systems-level approach. Previous studies have shown that acid production and tolerance in *S. mutans* requires large changes in gene expression and metabolic pathway utilization (3). For example, decreasing pH increases glycolytic activity and branched-chain amino acid synthesis without increasing cell growth (4). A drop in pH is also accompanied by an increased expression of F-ATPases to maintain a higher intracellular pH (5).

Mathematical models aid in our understanding of how an organism's genes collectively give rise to a phenotype. Models translate bioinformatic features (differential expression, presence/absence of genes) into biological function (flux distributions, uptake and secretion rates, and fitness). Constraint-based reconstruction and analysis (COBRA) of genome-scale models is widely used to integrate genetic and metabolic data to produce phenotypic predictions (6, 7). Models of microbial metabolism and transcriptional regulation predict responses to gene deletions (8, 9), mutation (10, 11),

metabolic shifts (12, 13), and long-term evolution (14, 15). The models identify emergent properties of a metabolic network, including links between pathways and inter-dependencies among genes (16, 17).

We present iSMU v1.0, a genome scale metabolic model for the *S. mutans* type strain UA159. Our model is manually curated using multiple databases, literature evidence, and phenotyping experiments. Our investigation of *S. mutans* focused our metabolism for two reasons: 1.) the primary metabolic products of *S. mutans*, lactic acid and biofilm matrix, are responsible for the pathogen's cariogenicity; and 2.) metabolic networks are among the best characterized intracellular networks with established computational techniques. Given metabolism's central role in cariogenesis, we believe the iSMU model will improve our understanding *S. mutans*' role in oral health.

Materials and Methods

Model Construction

The metabolic network of *S. mutans* UA159 was reconstructed following best practices in the COBRA modeling community (18). As summarized in Figure 1A, reconstruction began with the annotated UA159 genome (RefSeq GCA_000007465.2). Metabolic enzymes and the associated reactions were initially collected from KEGG (19) and Uniprot (20). The Metacyc (21), RHEA (22), ModelSEED (23), BiGG (24), and ChEBI (25) databases were used as secondary sources for metabolic reactions. Transport reactions were verified with TransportDB (26). When possible, KEGG identifiers were used for metabolites and reactions for consistency with other

databases. Custom identifiers (e.g. “add00001”) were added for reactions or metabolites without KEGG identifiers. Reactions without gene associations were only added when supported by experimental or literature evidence. These 11 reactions are explained in Table S1. All chemical species and formulas were converted to their protonation state at pH 7.0 using the ModelSEED database. A custom map of the iSMU model was constructed using Escher version 1.6.0 (Figure 2) (27).

Model Simulations using Flux Balance Analysis

To simulate growth using iSMU, reactions were collected into a stoichiometric matrix S where element $S(i,j)$ corresponds to the stoichiometric coefficient of the model’s i^{th} metabolite in the j^{th} reaction. $S(i,j)$ is negative if the metabolite is consumed and positive if the metabolite is produced. Two vectors of lower (l) and upper (u) bounds determine the reversibility of reactions. A vector of reaction fluxes v was calculated by maximizing the flux through the biomass reactions subject to mass balance constraints ($Sv = 0$) and reversibility constraints ($l \leq v \leq u$). To simulate gene deletions, the gene/protein/reaction rules for each reaction were evaluated to identify reactions that cannot carry flux in the deletion strain. The upper and lower bounds of these reaction were set to zero before maximizing flux through the biomass reaction. Genes were considered essential if their deletion allowed no biomass flux.

All simulations were performed with Matlab (version R2016b; MathWorks [https://www.mathworks.com]) using the COBRA toolbox (28). Mathematical programs were solved with Gurobi Optimizer (version 7.5; Gurobi Optimization

[http://www.gurobi.com]). Gene set enrichment for KEGG pathways was performed using DAVID (29, 30).

Model Availability

The final model is available as an SBML file and a spreadsheet compatible with the COBRA toolbox (Files S1 & S2). The model map is available as a JSON file (File S3) and SVG image (File S4). Future versions of the model and map will be available on the authors' website [http://jensenlab.net].

Strains and Media

S. mutans UA159 (ATCC 700610) was cultured on Brain-Heart Infusion (Sigma) agar plates or in Todd Hewitt broth with 0.3% yeast extract (Sigma). Strains were grown overnight in 5% CO_2 at 37□C unless specified.

Growth Assays

Growth experiments were performed in a Chemically Defined Medium (CDM) (31) with 22 amino acids, 11 vitamins, 3 nucleobases, 8 inorganic salts, and glucose (Table 1). Complete CDM or leave-one-out variants were prepared fresh weekly from concentrated stocks (31). All components were purchased from Sigma-Aldrich USA and were sterilized by autoclaving or filtration.

Overnight cultures of *S. mutans* were washed three times in sterile water. The overnight culture was concentrated 5x (from 5 ml to 1 ml), and 1 µl of the concentrate

was used to inoculate wells of a 96 well plate containing 200 µl of defined media. Plates were incubated without agitation in 5% CO_2 at 37□C. Optical density was measured by absorbance at 600 nm every hour for 16 h using a Tecan Infinite 200 Pro plate reader (Tecan, Männedorf, Switzerland). Exponential growth rates were calculated using the R package CellGrowth (version 3.7; Ludwig Maximilian University of Munich, [https://www.bioconductor.org/packages/release/bioc/html/cellGrowth.html]) with default settings. Growth rates were normalized to the growth rate in complete CDM.

Results

Manual curation produces an annotated metabolic model for S. mutans.

We manually reconstructed an *in silico* metabolic model for *S. mutans* type strain UA159 (Figure 1A). Our model, named iSMU for "*in silico S. mutans*", includes major metabolic pathways for carbohydrate metabolism and synthesis of amino acids, nucleotides, lipids, vitamins, and cofactors. The model includes 656 reactions transforming 514 metabolites. The reactions are catalyzed by the products of 488 genes (Figure 1B).

Our assembly of iSMU began with reaction databases and an annotated genome. Draft models assembled from genome annotations are often incomplete because of gaps in the genome annotation or spontaneous reactions that lack an associated enzyme. Several computational methods attempt to identify and add these missing reactions in a process called GapFilling (32). Rather than rely on automated GapFilling algorithms, we manually GapFilled iSMU by examining the reactions in each

pathway. We attempted to close gaps in any pathway that 1.) was complete except for a small number of reactions, or 2.) was blocked (unable to carry flux) due to metabolites that could not be produced or consumed. We also attempted to find unannotated or misannotated genes that could catalyze the GapFilled reactions. Compared to other metabolic reconstructions, our manually GapFilled model contains fewer incomplete pathways (Table S2 and Figure 1B). On average, published metabolic models of well-studied organisms lack gene annotations for 53% of the models' reactions (Table S2). These models also average 32.7% blocked reactions, i.e. reactions that cannot carry a steady state flux because they lack upsteam or downstream pathways. Our iSMU model has only 23.5% reactions without an associated enzyme and 2% blocked reactions.

Metabolic models simulate growth by collecting cellular building blocks into a biomass reaction. The biomass reaction is used as the objective function for metabolic simulations. A nonzero flux through the biomass reaction indicates growth in the metabolic environment specified by the model's inputs (called exchange reactions). We modified the biomass reaction from a model of *Enterococcus faecalis* V583 (33) to create a biomass reaction for *S. mutans* UA159. Both *E. faecalis* and *S. mutans* are lactic acid bacteria with similar metabolic capabilities. To tailor the biomass reaction to *S. mutans*, we changed the relative ratios of nucleotides and amino acids. We also changed the cell membrane composition to reflect membrane sugar polysaccharides specific to *S. mutans*. We replaced UDP-N-acetyl-D-galactosamine, which based on genetic evidence is not produced by *S. mutans*, with UDP-N-acetyl-D-mannosamine and UDP-N-acetyl-D-glucosamine. We also adjusted the cell wall fatty acids to their

measured proportions at pH 7.0 (34). The final biomass reaction consumes 56 metabolites to produce a unit of biomass.

The iSMU model reveals broad metabolic capabilities of S. mutans.

S. mutans can metabolize a wide range of carbon sources (35, 36). The ability to uptake and catabolize numerous carbohydrates allows *S. mutans* to thrive in the oral cavity of humans with varied diets. Besides glucose, the iSMU model can grow on fructose, sucrose, lactose, trehalose, ascorbate, arbutin, maltose, cellobiose, salicin, sorbitol, mannitol, mannose, N-acetyl glucosamine, fructan, galactose, galactinol, epimelibiose, melibiitol, melibiose, raffinose, maltodextrin, stachyose and malate. Growth on these carbohydrates is consistent with previous experimental studies (35, 36).

Carbohydrate metabolism is the dominant metabolic subsystem in *S. mutans* (37). Gene set enrichment for KEGG pathways classifies 134 (27.4%) of the genes in iSMU as carbohydrate metabolism (Figure 1C). By comparison, carbohydrate metabolism involves 20.7% of the genes in an *Escherichia coli* genome scale metabolic model (iJO1366, (38)) and 21.6% of the metabolic genes in a *Bacillus subtilis* genome scale model (iYO844, (39)) (Table S3).

Outside of carbohydrate metabolism, the largest subsystems in iSMU are: transport – 113 genes (23.2%); amino acid biosynthesis – 94 genes (19.3%); nucleotide biosynthesis – 56 genes (11.5%); lipid biosynthesis – 26 genes (5.3%); peptidoglycan biosynthesis – 23 genes (4.7%). (Figure 1C)

S. mutans has complex nutrient requirements for growth.

Several studies have investigated the minimal requirements for *S. mutans* growth *in vitro* (2, 40–43). Like many obligate human pathogens, *S. mutans* requires a combination of carbon, nitrogen, sulfur, and phosphorous sources; inorganic minerals; nucleotides; and vitamins and co-factors. We used our iSMU model and phenotypic assays to systematically explore auxotrophies for *S. mutans*.

The *S. mutans* UA159 genome encodes complete biosynthetic pathways for all 20 amino acids (37). *S. mutans* can grow without any exogenous amino acids using ammonium as the sole nitrogen source (40). The iSMU model can similarly produce biomass with ammonium and no amino acids. Using a series of leave-one-out experiments, we confirmed that the removal of individual amino acids from a defined media does not affect *S. mutans* growth in vitro (Figure 3). Simultaneous removal of cysteine and cystine does not significantly reduce growth, indicating that *S. mutans* can catabolize another sulfur source, possibly sulfate or methionine.

S. mutans can theoretically synthesize nucleotides (adenine, guanine, cytosine, uracil, and thymine) but only through the non-oxidative branch of the pentose phosphate pathway (44). *S. mutans* UA159 apparently lacks the more efficient oxidative branch of the pentose phosphate pathway (45). The non-oxidative pentose phosphate pathway is bidirectional and can produce or recycle ribose 5-phosphate and other pentose sugars. These sugars are necessary precursors for nucleotide biosynthesis. The model iSMU requires no nucleotides in the media for growth. However, we found that removing all

nucleotides from CDM prevents growth of *S. mutans* (Figure 3). We know that cytosine and thymine are not required for *S. mutans* growth since they are not present in CDM (Table 1). Consistent with model predictions, uracil and guanine can also be removed from CDM. Removing uracil does not significantly alter growth, but removing guanine causes a 30% decrease in growth rate (Figure 3). Only the removal of adenine completely abolished growth in CDM, which does not agree with our model predictions (Figure 3).

S. mutans is unable to synthesize thiamine, riboflavin, pyridoxal 5-phosphate, NAD^+/$NADP^+$, pantothenate, and folate. Anabolic pathways for these vitamins and co-factors are incomplete and key enzymes are not encoded in the *S. mutans* UA159 genome (Figure 2). All of these nutrients (or their metabolic precursors) are ingredients in two chemically defined media used to culture streptococci (CDM (31) and FMC (46)). Our model predicts aminobenzoate (an ingredient in CDM) can substitute for folate, but at least one of these nutrients is required for growth. Indeed, we found that *S. mutans* UA159 can grow in CDM without either folate or aminobenzoate but is unable to grow in media lacking both (Figure 3).

S. mutans cannot synthesize NAD^+/$NADP^+$ *de novo*. The iSMU model predicts that both NAD^+ and $NADP^+$ can be produced from any of NAD^+, $NADP^+$, nicotinamide, or nicotinate alone. CDM includes two of these four metabolites (NADP+ and nicotinamide). Consistent with our model, *S. mutans* can grow in CDM missing either NADP+ or nicotinamide, but not both (Figure 3).

Gene deletion simulations match experimental data.

Metabolic models contain chemical reactions and gene associations that link reactions to their corresponding enzymes. Gene associations are expressed as logical statements describing the required gene products for a reaction to carry flux. An enzymatic complex of two proteins is expressed using “and” (subunit 1 and subunit 2). A pair of isozymes that could each independently catalyze a reaction would be written with an “or” (isozyme 1 or isozyme 2). Flux balance analysis and the model’s gene associates can be combined to simulate the effects of gene deletions on growth. The logical rules in the gene associations are evaluated to identify reactions that cannot carry flux in a deletion strain. Reactions that cannot carry flux are removed from the model before calculating the maximum biomass flux. Deletion of an essential gene will prevent any nonzero biomass flux. Comparing experimentally determined essential genes with the model’s predictions to validate the model’s gene associations.

We simulated the effects of all single deletions for the 488 genes in the iSMU model. We compared the *in silico* deletions to two experimental gene deletion studies in *S. mutans* UA159: a transposon mutagenesis sequencing (Tn-seq) experiment (47) and a screen of an ordered array of single gene deletion strains (48). The Tn-seq study used a mariner-family transposon to generate random insertions across the UA159 genome (47). The transposon/genomic DNA junctions were amplified and sequenced to quantify fitness after growth in a defined media (FMC). Genes lacking transposon insertions sites are predicted to be essential in FMC. Overall, 84.8% of the essentiality predictions from the iSMU model were consistent with the Tn-seq data (Figure 4B). A comparison

between iSMU's essential gene predictions and the Tn-seq data is shown in the iSMU map in Figure 2.

The gene deletion strains in Quivey, et al., (48) were constructed individually using homologous recombination with a selective marker. The deletion library contains strains for 1112 of the 1956 genes in *S. mutans* UA159, including 366 of the 488 genes in the iSMU metabolic model. The remaining 122 genes are hypothesized to be essential or could not be deleted due to technical limitations. Deletion strains were grown in Brain Heart Infusion media (BHI), a rich and undefined media. We simulated BHI by opening all exchange reactions in the iSMU model. As shown in Figure 4A, 78.1% of the experimental essentiality results agreed with the model predictions. Agreement between iSMU's essential gene predictions and the data from Quivey, et. al. is highlighted the iSMU map in Figure S1.

Manual curation improves model consistency and accuracy.

Several software systems can generate draft metabolic models using reaction databases and annotated genomes. Our first attempt at reconstructing the metabolism of *S. mutans* used a draft model from the KBase system (49). Unfortunately, the draft model lacked many of the metabolic features of lactic acid bacteria and included several subsystems known to be inactive in homofermentive anaerobes. We therefore abandoned the KBase model and began a manual reconstruction process. We compared the final iSMU model to the KBase model to quantify the disagreement between the manual and automated reconstruction pipelines.

Our manually curated reconstruction differs substantially from the reconstruction produced automatically by KBase. Our iSMU model has 6.4% fewer reactions than the KBase model (656 vs. 701) but 97% more genes (488 vs. 247) (Figure 1B). Thus, the manually curated model has a larger proportion of gene associated reactions than the automated reconstruction. The dearth of gene associations in the KBase model is due in part to the 237 reactions added during GapFilling, since GapFilled reactions are added without genomic evidence for the reaction. By comparison, our iSMU model required only 11 GapFilled reactions to enable growth on defined media and 19 carbon sources.

The KBase model contains 41% more metabolites than our iSMU model. Unfortunately, 22 of the metabolites in the KBase model are “dead-end” metabolites that lack either a producing or consuming reaction. The dead-end metabolites block flux through 299 (43%) of the KBase model’s reactions. At steady-state, these blocked reactions cannot carry flux or be analyzed using Flux Balance Analysis. Less than 2% of the reactions in our iSMU model are blocked, indicating more complete reaction pathways than the automated reconstruction.

Discussion

iSMU is the first whole genome metabolic model of the cariogenic pathogen *S. mutans* UA159. The model captures the entire metabolism of the organism and was validated by comparing model predictions to experimental evidence. Metabolism plays a dual role in the pathogenicity of *S. mutans*. First, fermenting sugars creates caries-

causing lactic acid. A significant portion of the *S. mutans* genome is dedicated to carbohydrate metabolism, reflecting the plasticity of *S. mutans*' metabolism. Second, multiple metabolic subsystems are required for *S. mutans* to tolerate acid and outcompete non-cariogenic streptococci. A mathematical model allows us to investigate connections among metabolic pathways during pathogenesis.

iSMU's predictions agree with most of the nutrient depletion experiments, but some of *S. mutans*' auxotrophies are unexplained by the model. For example, the UA159 genome encodes a complete pathway for adenine synthesis, but exogenous adenine is required for growth *in vitro*. The adenine synthesis pathway in iSMU may not be expressed or functional in UA159 when grown aerobically in CDM. Other experiments agree qualitatively, but not quantitatively with the model. When guanine, aminobenzoate, nicotinamide, or sodium acetate are removed from CDM, S. mutans grows slower predicted by the model. The model also underpredicts growth rates when pyridoxal and pyridoxamine, riboflavin, and thamine are removed. Differences like these are expected with constraint-based models that lack kinetic details for nutrient uptake and enzymatic turnover.

S. mutans UA159 can grow on minimal media with ammonium as the sole nitrogen source (40), and the iSMU model can produce biomass in these conditions. Experimentally, growth on ammonium requires an anaerobic environment, but the model can produce biomass with or without oxygen. Oxygen may repress expression of enzymes required for scavenging nitrogen from ammonium, and the lack of regulation in our model would explain why iSMU can grow aerobically using ammonium.

Several factors could explain the disagreements between the model's essentiality prediction and experimental results. First, we note that both methods for identifying essential genes have technical strengths and limitations. In the ordered gene deletion library, any gene for which a deletion mutant cannot be constructed is labeled essential. A nonessential gene located in a region of the chromosome refractory to homologous recombination would be incorrectly labeled as essential. The Tn-seq libraries in Shields, et al. were constructed using *in vitro* transposon mutagenesis followed by homologous recombination, so the same limitation applies. The Tn-seq libraries were grown for ~30 generations before sequencing. Such a large expansion can bias the library against mutants with large fitness defects. Although these mutants may be viable, they appear at such low frequency in the final pool that they are missed during sequencing. The corresponding genes would be incorrectly labeled as essential. The disagreement between the Tn-seq and defined deletion library suggest the "essential genome" of *S. mutans* UA159 has not been fully elucidated.

Inaccuracies in the *iSMU* model also contribute to disagreements over essential genes. After decades of curation, metabolic models for the model organisms *E. coli* and *S. cerevisiae* still miss some essential gene predictions (38, 50). Unannotated genes could catalyze redundant routes to synthesize essential metabolites *in vivo*, creating false positive essentiality predictions in iSMU. Regulation, loss of function mutations, and missing cofactors can also restrict the metabolic capabilities of *S. mutans*, making the pathogen less metabolically flexible than the iSMU model.

Overall, believe the model's predictions could be improved by either 1.) incorporating regulatory rules during simulations, or 2.) using gene expression or other high-throughput data to tailor the model to anaerobic, aerobic, acidic, or other conditions. *S. mutans*' metabolic requirements change as the bacterium encounters different niches in the mouth. Before forming thick biofilms and deep dental caries, growth conditions are likely aerobic with abundant nutrients from saliva and food consumed by the host. Deep dental caries may create anaerobic conditions with limited nutrient availability. In this environment, *S. mutans* would need to synthesize many biomass components de novo.

S. mutans is a model organism in oral microbiology (51). Our iSMU model draws from hundreds of studies to form an accurate, genome-wide picture of *S. mutans* metabolism. The model also highlights the value added by manual curation. The metabolism of *S. mutans* is well characterized on the molecular and pathway levels. Incorporating manually curated models of *S. mutans* and other lactic acid bacteria may improve the accuracy of automatic reconstruction pipelines.

Acknowledgements

We thank Robert Shields for assistance with the Tn-seq data and John Gerlt for sharing equipment and reagents. This work was supported by the NIH National Institute of Dental and Craniofacial Research (grant DE026817 to PAJ) and the University of Illinois at Urbana-Champaign. The authors declare no financial conflict of interest.

References

1. Daneo-Moore L, Terleckyj B, Shockman GD. 1975. Analysis of growth rate in sucrose-supplemented cultures of *Streptococcus mutans*. Infect Immun 12:1195–1205.

2. Terleckyj B, Shockman GD. 1975. Amino acid requirements of *Streptococcus mutans* and other oral streptococci. Infect Immun 11:656–664.

3. Baker JL, Abranches J, Faustoferri RC, Hubbard CJ, Lemos JA, Courtney MA, Quivey R. Transcriptional profile of glucose-shocked and acid-adapted strains of *Streptococcus mutans*. Molecular Oral Microbiology 30:496–517.

4. Hamilton IR, Ellwood DC. 1978. Effects of Fluoride on Carbohydrate Metabolism by Washed Cells of *Streptococcus mutans* Grown at Various pH Values in a Chemostat. Infect Immun 19:434–442.

5. Kuhnert WL, Zheng G, Faustoferri RC, Quivey RG. 2004. The F-ATPase Operon Promoter of *Streptococcus mutans* Is Transcriptionally Regulated in Response to External pH. J Bacteriol 186:8524–8528.

6. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. Molecular Systems Biology 5:320.

7. R. Hyduke D, E. Lewis N, Ø. Palsson B. 2013. Analysis of omics data with genome-scale models of metabolism. Molecular BioSystems 9:167–174.

8. Suthers PF, Zomorrodi A, Maranas CD. Genome-scale gene/reaction essentiality and synthetic lethality analysis. Molecular Systems Biology 5:301.

9. Shlomi T, Berkman O, Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. PNAS 102:7695–7700.

10. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, Adkins JN, Schramm G, Purvine SO, Lopez-Ferrer D, Weitz KK, Eils R, König R, Smith RD, Palsson BØ. 2010. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. Molecular Systems Biology 6:390.

11. Applebee MK, Herrgård MJ, Palsson BØ. 2008. Impact of Individual Mutations on Increased Fitness in Adaptively Evolved Strains of *Escherichia coli*. J Bacteriol 190:5087–5094.

12. Jensen PA, Papin JA. 2011. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 27:541–547.

13. Vu TT, Stolyar SM, Pinchuk GE, Hill EA, Kucek LA, Brown RN, Lipton MS, Osterman A, Fredrickson JK, Konopka AE, Beliaev AS, Reed JL. 2012. Genome-Scale Modeling of Light-Driven Reductant Partitioning and Carbon Fluxes in Diazotrophic Unicellular Cyanobacterium *Cyanothece* sp. ATCC 51142. PLOS Computational Biology 8:e1002460.

14. Fong SS, Palsson BØ. 2004. Metabolic gene–deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nature Genetics 36:1056–1058.

15. Ibarra RU, Edwards JS, Palsson BO. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. Nature 420:186–189.

16. Bilu Y, Shlomi T, Barkai N, Ruppin E. 2006. Conservation of Expression and Sequence of Metabolic Genes Is Reflected by Activity Across Metabolic States. PLOS Computational Biology 2:e106.

17. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD. 2004. Flux Coupling Analysis of Genome-Scale Metabolic Network Reconstructions. Genome Res 14:301–312.

18. Thiele I, Palsson BØ. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5:93–121.

19. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361.

20. Bateman A, Martin MJ, O'Donovan C, Magrane M, Alpi E, Antunes R, Bely B, Bingley M, Bonilla C, Britto R, Bursteinas B, Bye-A-Jee H, Cowley A, Silva AD, Giorgi MD, Dogan T, Fazzini F, Castro LG, Figueira L, Garmiri P, Georghiou G, Gonzalez D, Hatton-Ellis E, Li W, Liu W, Lopez R, Luo J, Lussi Y, MacDougall A, Nightingale A, Palka B, Pichler K, Poggioli D, Pundir S, Pureza L, Qi G, Renaux A, Rosanoff S, Saidi R, Sawford T, Shypitsyna A, Speretta E, Turner E, Tyagi N, Volynkin V, Wardell T, Warner K, Watkins X, Zaru R, Zellner H, Xenarios I, Bougueleret L, Bridge A, Poux S, Redaschi N, Aimo L, Argoud-Puy G, Auchincloss A, Axelsen K, Bansal P, Baratin D, Blatter M-C, Boeckmann B, Bolleman J, Boutet E, Breuza L, Casal-Casas C, Castro E de, Coudert E, Cuche B, Doche M, Dornevil

D, Duvaud S, Estreicher A, Famiglietti L, Feuermann M, Gasteiger E, Gehant S, Gerritsen V, Gos A, Gruaz-Gumowski N, Hinz U, Hulo C, Jungo F, Keller G, Lara V, Lemercier P, Lieberherr D, Lombardot T, Martin X, Masson P, Morgat A, Neto T, Nouspikel N, Paesano S, Pedruzzi I, Pilbout S, Pozzato M, Pruess M, Rivoire C, Roechert B, Schneider M, Sigrist C, Sonesson K, Staehli S, Stutz A, Sundaram S, Tognolli M, Verbregue L, Veuthey A-L, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Garavelli JS, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Yeh L-S, Zhang J. 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169.

21. Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD. 2018. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res 46:D633–D639.

22. Morgat A, Lombardot T, Axelsen KB, Aimo L, Niknejad A, Hyka-Nouspikel N, Coudert E, Pozzato M, Pagni M, Moretti S, Rosanoff S, Onwubiko J, Bougueleret L, Xenarios I, Redaschi N, Bridge A. 2017. Updates in Rhea – an expert curated resource of biochemical reactions. Nucleic Acids Res 45:D415–D418.

23. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. Nature Biotechnology 28:977–982.

24. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, Ebrahim A, Palsson BO, Lewis NE. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res 44:D515–D522.

25. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, Turner S, Swainston N, Mendes P, Steinbeck C. 2016. ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res 44:D1214–D1219.

26. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. 2017. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic Acids Res 45:D320–D324.

27. King ZA, Dräger A, Ebrahim A, Sonnenschein N, Lewis NE, Palsson BO. 2015. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. PLOS Computational Biology 11:e1004321.

28. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V, Magnusdóttir S, Ng CY, Preciat G, Žagare A, Chan SHJ, Aurich MK, Clancy CM, Modamio J, Sauls JT, Noronha A, Bordbar A, Cousins B, Assal DCE, Valcarcel LV, Apaolaza I, Ghaderi S, Ahookhosh M, Guebila MB, Kostromins A, Sompairac N, Le HM, Ma D, Sun Y, Wang L, Yurkovich JT, Oliveira MAP, Vuong PT, Assal LPE, Kuperstein I, Zinovyev A, Hinton HS, Bryant WA, Artacho FJA, Planes FJ, Stalidzans E, Maass A, Vempala S, Hucka M, Saunders MA, Maranas CD, Lewis NE, Sauter T, Palsson BØ, Thiele I, Fleming RMT. 2017. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3.0. arXiv:171004038.

29. Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols 4:44–57.

30. Huang DW, Sherman BT, Lempicki RA. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1–13.

31. Chang JC, LaSarre B, Jimenez JC, Aggarwal C, Federle MJ. 2011. Two Group A Streptococcal Peptide Pheromones Act through Opposing Rgg Regulators to Control Biofilm Development. PLOS Pathogens 7:e1002190.

32. Pan S, Reed JL. 2018. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. Current Opinion in Biotechnology 51:103–108.

33. Veith N, Solheim M, Grinsven KWA van, Olivier BG, Levering J, Grosseholz R, Hugenholtz J, Holo H, Nes I, Teusink B, Kummer U. 2015. Using a Genome-Scale Metabolic Model of *Enterococcus faecalis* V583 To Assess Amino Acid Uptake and Its Impact on Central Metabolism. Appl Environ Microbiol 81:1622–1633.

34. Quivey RG, Faustoferri R, Monahan K, Marquis R. 2000. Shifts in membrane fatty acid profiles associated with acid adaptation of *Streptococcus mutans*. FEMS Microbiol Lett 189:89–92.

35. Ikeda T, Sandham HJ. 1972. A high-sucrose medium for the identification of *Streptococcus mutans*. Arch Oral Biol 17:781–783.

36. Moye ZD, Burne RA, Zeng L. 2014. Uptake and Metabolism of N-Acetylglucosamine and Glucosamine by *Streptococcus mutans*. Appl Environ Microbiol 80:5053–5067.

37. Ajdić D, McShan WM, McLaughlin RE, Savić G, Chang J, Carson MB, Primeaux C, Tian R, Kenton S, Jia H, Lin S, Qian Y, Li S, Zhu H, Najar F, Lai H, White J, Roe BA, Ferretti JJ. 2002. Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental pathogen. PNAS 99:14434–14439.

38. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BØ. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Molecular Systems Biology 7:535.

39. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R. 2007. Genome-scale Reconstruction of Metabolic Network in *Bacillus subtilis* Based on High-throughput Phenotyping and Gene Essentiality Data. J Biol Chem 282:28791–28799.

40. Martin EJS, Wittenberger CL. 1980. Regulation and function of ammonia-assimilating enzymes in *Streptococcus mutans*. Infect Immun 28:220–224.

41. Cowman RA, Perrella MM, Fitzgerald RJ. 1974. Influence of Incubation Atmosphere on Growth and Amino Acid Requirements of *Streptococcus mutans*. Appl Environ Microbiol 27:86–92.

42. Griffith CJ, Carlsson J. 1974. Mechanism of Ammonia Assimilation in Streptococci. Microbiology 82:253–260.

43. Carlsson J. 1970. Nutritional Requirements of *Streptococcus mutans*. CRE 4:305–320.

44. Bridges RB. 1977. Ribose biosynthesis in *Streptococcus mutans*. Archives of Oral Biology 22:139–145.

45. Richardson AR, Somerville GA, Sonenshein AL. 2015. Regulating the Intersection of Metabolism and Pathogenesis in Gram-positive Bacteria. Microbiology Spectrum 3.

46. Terleckyj B, Willett NP, Shockman GD. 1975. Growth of several cariogenic strains of oral streptococci in a chemically defined medium. Infect Immun 11:649–655.

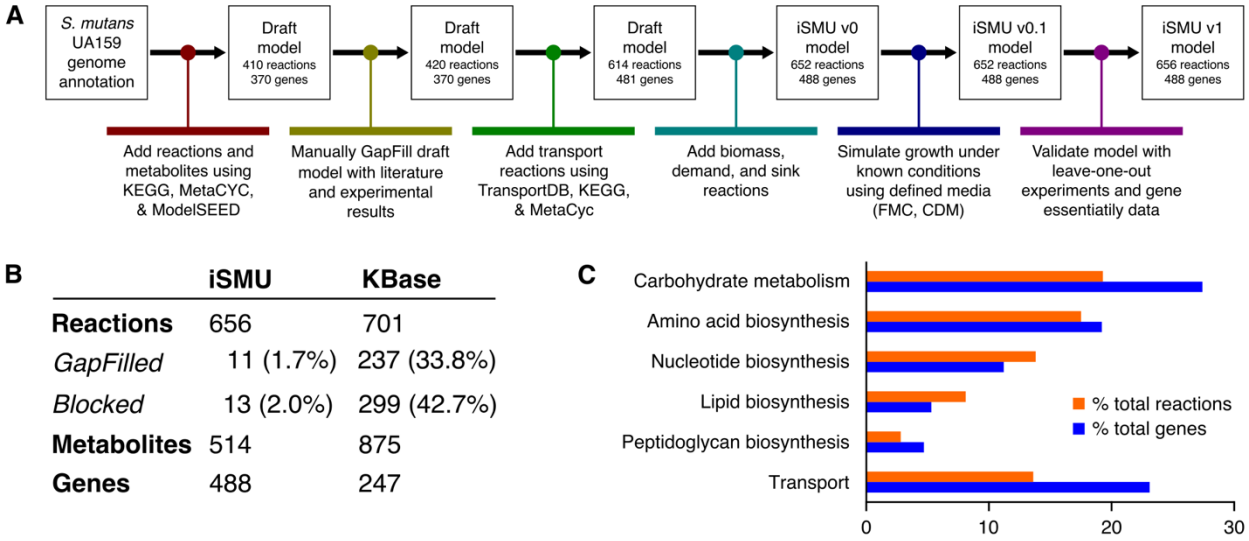
47. Shields RC, Zeng L, Culp DJ, Burne RA. 2018. Genomewide Identification of Essential Genes and Fitness Determinants of *Streptococcus mutans* UA159. mSphere 3:e00031-18.

48. Quivey RG, Grayhack EJ, Faustoferri RC, Hubbard CJ, Baldeck JD, Wolf AS, MacGilvray ME, Rosalen PL, Scott-Anne K, Santiago B, Gopal S, Payne J, Marquis RE. 2015. Functional profiling in *Streptococcus mutans*: construction and examination of a genomic collection of gene deletion mutants. Molecular Oral Microbiology 30:474–495.

49. Arkin AP, Stevens RL, Cottingham RW, Maslov S, Henry CS, Dehal P, Ware D, Perez F, Harris NL, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Gunter D, Murphy-Olson D, Chan S, Kamimura RT, Brettin TS, Meyer F, Chivian D, Weston DJ, Glass EM, Davison BH, Kumari S, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, DeJongh M, Devoid S, Dietrich E, Drake MM, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Mills M, Novichkov P, Oh T, Olsen GJ, Olson B, Parrello B, Pasternak S, Pearson E, Poon SS, Price G, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S. 2016. The DOE Systems Biology Knowledgebase (KBase). bioRxiv 096354.

50. Zomorrodi AR, Maranas CD. 2010. Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. BMC Systems Biology 4:178.

51. Lemos JA, Quivey RG, Koo H, Abranches J. 2013. *Streptococcus mutans*: a new Gram-positive paradigm? Microbiology 159:436–445.



B

	iSMU	KBase
Reactions	656	701
GapFilled	11 (1.7%)	237 (33.8%)
Blocked	13 (2.0%)	299 (42.7%)
Metabolites	514	875
Genes	488	247

Figure 1: A. Reconstruction of the *S. mutans* metabolic network began with an annotated UA159 genome. The draft model was refined with bioinformatics databases and experimental results. **B.** The manually reconstructed model (iSMU) has fewer GapFilled (non-gene associated) and blocked reactions than a model generated automatically by the KBase systems. **C.** The iSMU model contains reactions across multiple KEGG pathways, especially carbohydrate catabolism and transport.

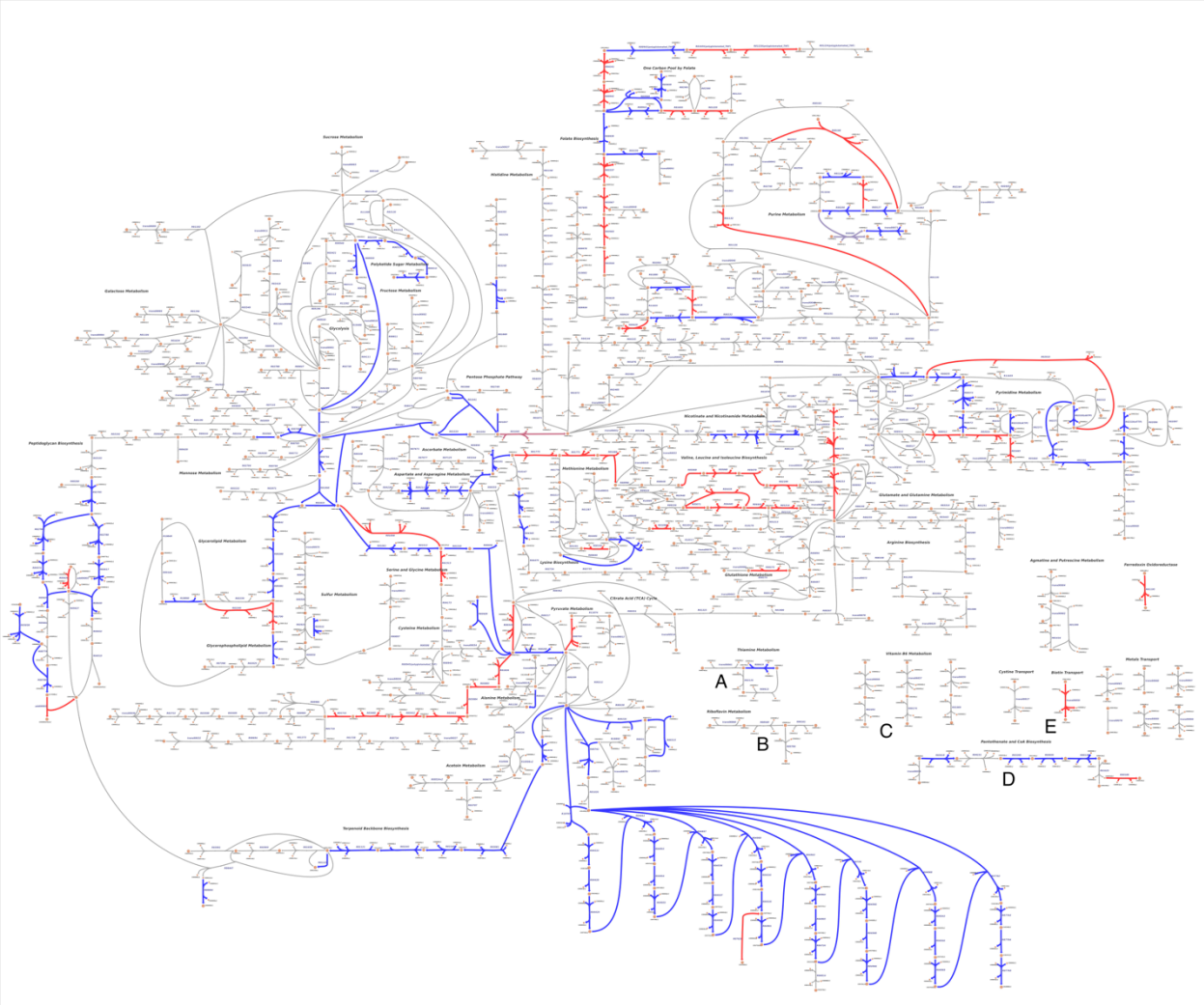


Figure 2: A custom pathway map shows all reactions in the iSMU model. (A high-resolution image is available as File S4.) *S. mutans* UA159 appears to lack complete pathways for synthesizing thiamine (**A**), riboflavin (**B**), pyridoxine (**C**), pantothenate (**D**), and biotin (**E**). Reactions are colored by agreement between the essentiality predictions of the associated genes and Tn-seq data from Shields et al. [ref]. Blue reactions agree with the Tn-seq essentiality results; red reactions disagree. Overall agreement between the datasets is 84.8% (see Figure 4).

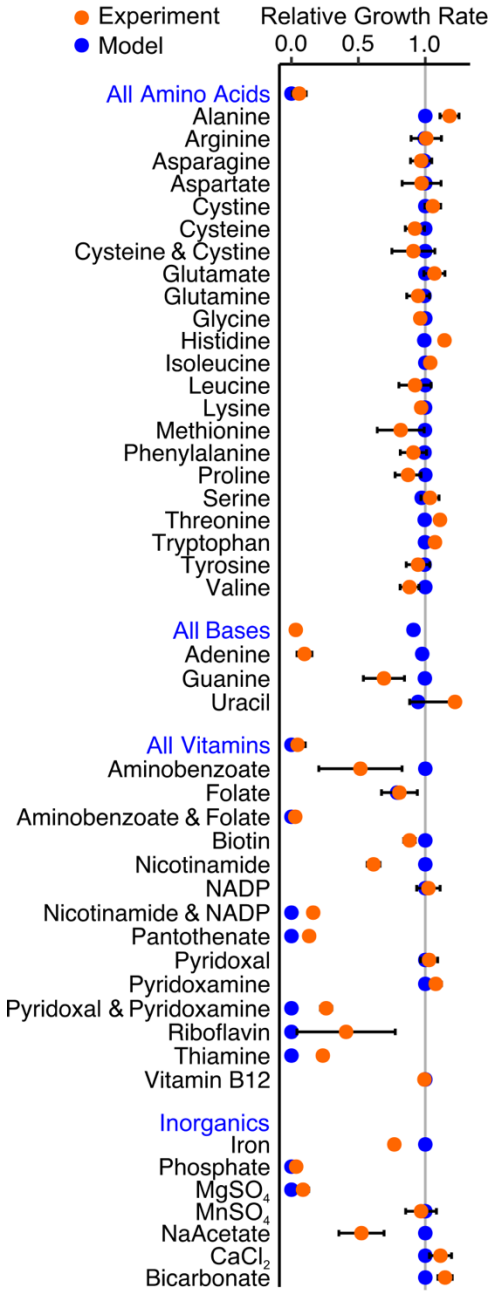


Figure 3: Model predictions (blue) match growth experiments for *S. mutans* UA159 in defined media (orange). Growth rate was measured for CDM lacking the specified component(s). Blue labels indicate removal of all components listed below. Experimental data are means of three independent trials with error bars representing the standard deviation. Growth rates are normalized to *S. mutans* UA159 grown in complete CDM.

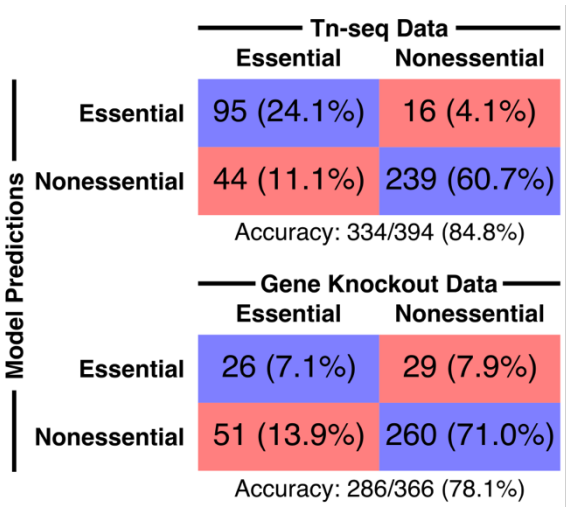


Figure 4: iSMU essentiality predictions align with two experimental studies. Shields et al. (47) (top) used transposon mutagenesis sequencing (Tn-seq) to identify essential genes in the defined media FMC. Quivey, et al. (48) (bottom) screened a library of single gene knockout strains for growth in rich media. Blue boxes indicate the number (percentage) of genes in the model and dataset that are both essential or nonessential. Red boxes indicated disagreements between the model and experiments.

Table 1: CDM is composed of 22 amino acids, 11 vitamins, 3 nucleobases, 8 inorganic salts, and glucose.

Component	Concentration [g/L]
deionized H_2O	-
iron	0.006
phosphate	18.3
$MgSO_4 \cdot 7H_2O$	0.7
$MnSO_4 \cdot H_2O$	0.005
$NaC_2H_3O_2 \cdot 3H_2O$	4.5
DL-alanine	0.1
L-arginine	0.1
L-aspartic acid	0.1
L-asparagine	0.1
L-cysteine HCl	0.65
L-cystine	0.05
L-glutamic acid	0.1
L-glutamine	0.2
glycine	0.1
L-histidine	0.1
L-isoleucine	0.1
L-leucine	0.1
L-lysine	0.1
L-methionine	0.1
L-phenylalanine	0.1
L-proline	0.1
hydroxy-L-proline	0.1
L-serine	0.1
L-threonine	0.2
L-tryptophan	0.1
L-tyrosine	0.1
L-valine	0.1
P-aminobenzoic acid	0.0002
biotin	0.0002
folic acid	0.0008
nicotinamide	0.001
B-nicotinamide adenine dinucleotide phosphate	0.0025
pantothenate calcium salt	0.002
pyridoxal	0.001
pyridoxamine dihydrochloride	0.001
riboflavin	0.002
thiamine hydrochloride	0.001
vitamin B12	0.0001
adenine	0.02
guanine hydrochloride	0.02
uracil	0.02
$CaCl_2$	0.00676
$NaHCO_3$	2.5
glucose	10

Supplemental Material

Table S1: GapFilled reactions in iSMU and evidence for inclusion.

Table S2: Curated genome-scale metabolic models used to compare number of reactions without a gene association and blocked reactions.

Table S3: Model comparisons between iSMU, iJO1366, and iYO844.

File S1: iSMU genome-scale metabolic model in SBML file format.

File S2: iSMU genome-scale metabolic model in spreadsheet file format.

File S3: iSMU's entire model map in JSON file format.

File S4: iSMU's entire model map in SVG file format.