Metabolite Damage and Damage-Control in a Minimal Genome

Haas, Drago; Thamm, Antje M.; Sun, Jiayi; Huang, Lili; Sun, Lijie; Beaudoin, Guillame A.W.; Wise, Kim S.; Lema-Ortiz, Claudia; Bruner, Steven D.; Breuer, Marian; Luthey-Schutten, Zaida; Lin, Jushen; Wilson, Mark A.; Brown, Greg; Yakunin, Alexander; Kurilyak, Inna; Folz, Jacob; Fiehn, Oliver; Glass, John I.; Hanson, Andrew D.; Henry, Christopher S.; de Crecy-Lagard, Valerie

mBio

DOI:
10.1128/mbio.01630-22

Published: 30/08/2022

Peer reviewed version

Metabolite Damage and Damage-Control in a Minimal Genome

Drago Haas[1&], Antje M. Thamm[2%], Jiayi Sun[2$], Lili Huang[3#], Lijie Sun[4], Guillaume A.W. Beaudoin[2†], Kim S. Wise[4], Claudia Lerma-Ortiz[2], Steven D. Bruner[5], Marian Breuer[6], Zaida Luthey-Schulten[7], Jiusheng Lin[8], Mark A. Wilson[8], Greg Brown[9,10], Alexander F. Yakunin[9], Inna Kurilyak[11], Jacob Folz[11], Oliver Fiehn[11], John I. Glass[4], Andrew D. Hanson[2], Christopher S. Henry[12,13*] and Valérie de Crécy-Lagard[1,14 *]

[1] Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA

[2] Horticultural Sciences Department, University of Florida, Gainesville, FL 3261, USA

[3] Food Science and Human Nutrition Department, University of Florida, Gainesville, FL 32611, USA

[4] J. Craig Venter Institute, La Jolla, California USA

[5] Chemistry Department, University of Florida, Gainesville, Florida, USA

[6] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands

[7] Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[8] Department of Biochemistry and the Redox Biology Center, University of Nebraska, Lincoln, Nebraska, USA

[9] Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Canada

[10] Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, UK

[11] West Coast Metabolomics Center, UC Davis, Davis, California, USA

[12] Data Science and Learning, Argonne National Laboratory, Argonne, Illinois, USA

[13] Consortium for Advanced Science and Engineering, The University of Chicago, Chicago, Illinois, USA

[14] University of Florida Genetics Institute, Gainesville, Florida, USA11

[&] Current address – Sanofi, 13 Quai Jules Guesde, Vitry-sur-Seine, France

[%] Current address – Havas Life Bird and Schulte, Urachstrasse 19, Freiburg im Breisgau, Germany

[$] Current address – Captozyme, 1622 NW 55th Place, Gainesville, Florida, USA

[#] Current address – Lingnan Medical Research Center, Guangzhou University of Chinese Medicine, Guangzhou, Guangdong, China,

† Current address –Ginkgo Bioworks, 27 Drydock Ave 8th Floor, Boston, Massachussets, USA

* Corresponding Authors: Valérie de Crécy-Lagard vcrecy@ufl.edu and Christopher S. Henry chenry@anl.gov

Abstract

Analysis of the genes retained in the minimized Mycoplasma JCVI-Syn3A genome established that systems that repair or preempt metabolite damage are essential to life. Several genes known to have such functions were identified and experimentally validated, including 5-formyltetrahydrofolate cyclo-ligase, CoA disulfide reductase, and certain hydrolases. Furthermore, we discovered that an enigmatic YqeK hydrolase domain fused to NadD has a novel proofreading function in NAD synthesis and could double as a MutT-like sanitizing enzyme for the nucleotide pool. Finally, we combined metabolomics and cheminformatics approaches to extend the core metabolic map of JCVI-Syn3A to include promiscuous enzymatic reactions and spontaneous side reactions. This extension revealed that several key metabolite damage-control systems remain to be identified in JCVI-Syn3A, such as that for methylglyoxal.

Importance

Metabolite damage and repair mechanisms are being increasingly recognized. We present here compelling genetic and biochemical evidence for the universal importance of these mechanisms by demonstrating that stripping a genome down to its barest essentials leaves metabolite damage-control systems in place. Furthermore, our metabolomic and cheminformatic results point to the existence of a network of metabolite damage and damage-control reactions that extends far beyond the corners of it that have been characterized so far. In sum, there can be little room left to doubt that metabolite damage and the systems that counter it are mainstream metabolic processes that cannot be separated from life itself.

Introduction

A foundational goal of synthetic biology was to create a minimal living organism by a bottom-up approach (1). This goal was reached in 2016 with the creation of JCVI-Syn3.0 (2). This organism was built from the ruminant pathogen *Mycoplasma mycoides capri* serovar LC GM12 by DNA synthesis, recombination, and genome transplantation techniques, and included only genes required for survival or to support a reasonable growth rate (428 protein-coding genes and 34 RNA genes) (2). The initial JCVI-Syn3.0 strain was extremely fragile; a derivative with 18 more genes, JCVI-Syn3A was more stable and was the basis for a metabolic model (3). Surprisingly, when the JCVI-Syn3.0 was published in 2016, ~30 % of its genes could not be assigned a specific function. The initial annotation has since been improved by manual curation (4), metabolic modeling (3), and further in silico analyses (5) but ~85 proteins with unknown or vaguely defined functions remain (Supplemental data A1). These unknowns cannot all be missing parts of synthesis/breakdown pathways as the metabolic reconstruction identified only four metabolic and eight transport reactions as missing (3).

A crucial area of metabolism usually left out of metabolic models is metabolite damage and repair. Enzymes make mistakes and metabolites undergo spontaneous chemical reactions (6, 7). These damage reactions are ever-present and, when the resulting products are toxic, can reduce fitness (6, 8). It has been shown recently that many enzymes of formerly unknown function repair or pre-empt metabolite damage (9–11), that mutations in metabolite repair enzymes cause human diseases (12–14), and that pathway engineering can fail unless appropriate repair enzymes are installed (15). The emerging recognition of the nature and extent of metabolite damage and repair raised the question of the importance of metabolite repair for a minimal genome like JCVI-Syn3/3A. By combining expert manual curation, comparative genomics, metabolomics, metabolic modeling, cheminformatics, and experimental validation, we identified a set of chemical damage reactions likely to occur in JCVI-Syn3 and some of the damage repair and preemption activities that this minimal genome encodes.

Results and Discussion

Identification and validation of homologs of known metabolite repair enzymes

We first manually screened the predicted proteome of JCVI-Syn3A for homologs of known metabolite repair enzymes (6, 15, 16) (see Supplemental data S1 and Appendix). Several were found, as follows.

1. 5-FCL. 5-Formyltetrahydrofolate (5-CHO-THF) is a by-product of serine hydroxymethyltransferase (SHMT) (17)(Fig. 1A) that inhibits folate-dependent enzymes and must therefore be recycled or destroyed (18). Of various enzymes known to recycle 5-CHO-THF (19), the most widespread is 5-formyltetrahydrofolate cyclo-ligase (5-FCL) (encoded by *fau/ygfA* (16) in *E. coli*). The JCVI-syn3A genome encodes a 5-FCL homolog (JCVISYN3A_0443); this gene was confirmed to encode an active 5-FCL by a complementation assay. Specifically, an *E. coli* K12 Δ*ygfA* strain does not grow on M9 minimal medium with 0.2% glucose as carbon source and 20 mM glycine as sole nitrogen source (19) (Fig. 1B). Expression of JCVISYN3A_0443 from a plasmid complemented this growth phenotype (Fig. 1B). Note that the essentiality of JCVISYN3A_0443 might be due both to its repair function and to a role as a source of 5,10-methenyltetrahydrofolate-polyglutamate (3).

2. Thiol reductases. Like all aerobes, JCVI-syn3A encounters oxidative stress that can damage macromolecules. Maintaining protein and small-molecule thiol groups in their reduced state is critical for cellular redox homeostasis (20). Thioredoxin/thioredoxin reductase is the dominant protein thiol oxidoreductase system in many organisms, using reducing equivalents ultimately derived from NAPDH (21, 22). The JCVI-Syn3A genome encodes homologs of the thioredoxin system proteins (TrxB/JCVISYN3A_0819 and TrxA/JCVISYN3A_0065) that are most likely involved in reducing protein disulfide bonds and have been partially characterized in other *Mycoplasma* species (Fig. 2A)(23, 24). Both genes are essential (Supplemental data A1), supporting key roles for TrxA and TrxB in disulfide bond reduction. Note, however, that thioredoxin is also the electron donor for ribonucleotide reductase, so that JCVISYN3A_0819 and JCVISYN3A_0065 may be essential for this reason (23, 25).

JCVI3_0887 is a homolog of CoA disulfide reductase (CoADR), which may have a major redox role in certain bacteria (26). Because CoA is required for several reactions in the JCVI-syn3A metabolic model and is predicted to be imported from the medium, CoADR could maintain the CoA pool in the reduced state. Testing the CoADR activity of the JCVISYN3A_0887 showed that it is an active CoAD reductase that operates well at physiological pH (pH 7.5) (27) and has reasonable K_M (0.17 mM) and k_{cat} (2.8 s^{-1}) values (Fig.

2B). It lacks detectable activity against oxidized glutathione or pantethine (Fig. 2C). While we cannot exclude the possibility that reduced glutathione is imported from the medium and oxidized glutathione is exported, a CoA-based system is a more parsimonious solution to the redox balance problem.

Functional analysis of HAD proteins identifies a nucleotide phosphatase with possible dual roles

Our second strategy to identify metabolite repair enzymes was based on the demonstration that hydrolases of previously uncertain or unknown function were subsequently shown to participate in metabolite repair (9). Five genes encoding stand-alone members of the HAD (haloacid dehalogenase) hydrolase family (28) were identified in the JCVI-Syn3A genome (Supplemental data S1) and are conserved in closely related *Mesoplasma florum* L1 genome (29) (Table 1). Such HAD hydrolases often participate in metabolite repair or homeostasis, as many damaged or toxic intermediates are phosphorylated (e.g. phosphosugars), and their recycling or removal requires a phosphatase (9, 30).

Comparative genomic analysis of the stand-alone HADs did not point to clear functional hypotheses, except for JCVISYN3A_0728, whose location in a predicted operon with triose-phosphate isomerase and phosphoglycerate mutase suggested a role in sugar phosphate metabolism (Table 1). Possible functions for the HAD proteins included: 1) repair of substrates to be identified; 2) missing phosphatases involved in primary metabolism identified by the metabolic model such as sedoheptulose 1,7-bisphosphate phosphatase or phosphatidate phosphatase; 3) nucleotide phosphatases involved in dNTP pool maintenance. To discriminate among these hypotheses, we combined biochemistry, genetics, and metabolomics.

The four HAD proteins that we were able to express in *E. coli* (JCVISYN3A_0066, JCVISYN3A_0077, JCVISYN3A_0728, JCVISYN3A_0907) were tested for activity against a panel of 94 phosphatase substrates (Table A1) (31). The four proteins had detectable activity against the model phosphatase substrate *p*-nitrophenyl phosphate (*p*NPP) and different physiological substrates (Fig. A1). The JCVISYN3A_0728 enzyme hydrolyzed a wide range of nucleoside and sugar phosphates, the JCVISYN3A_0907 and JCVISYN3A_0077 enzymes hydrolyzed narrower ranges of sugar phosphates, and the JCVISYN3A_0066 enzyme hydrolyzed FMN and CoA. That sugar phosphates are good substrates of JCVISYN3A_0728 is

consistent with its genomically-predicted role in sugar phosphate metabolism, but no specific function or substrate could be assigned. Note, however, that the 94-substrate panel did not include damaged sugar phosphates.

We attempted to delete HAD-encoding genes in JCVI-syn3A,expecting this to be possible because transposon bombardment of the JCVI-syn3A genome indicated all five HADs were quasi-essential (i.e., required for fast growth but not for viability) [(3) and Supplemental data S1]. Deletants were readily obtained for genes JCVISYN3A_0066, JCVISYN3A_0077, JCVISYN3A_0728, and JCVISYN3A_0907 (Supplemental data S2). Attempts to delete JCVISYN3A_0710 using two different methods were unsuccessful (Supplemental data S2). Deletion of this gene could have resulted in an extremely slow-growing strain that was unrecoverable in the conditions used, Alternatively JCVISVN3A_0710 could be essential, the transposon insertions in the gene being artifacts. That the same gene is also essential in *M. florum* (Table 1*)* favors the latter hypothesis.

We observed no major differences in growth rates between JCVI_Syn3A and any HAD mutant (Fig. A2). To conduct a metabolomics analysis, the four mutants and the JCVI-Syn3A parent were grown in SP4-KO medium and harvested at the same point of log-phase growth. (Appendix and Tables A2 and A3). A total of 4152 features were detected in the samples using hydrophilic interaction liquid chromatography (HILIC) and mass spectrometry (Supplemental data S3), of which 522 were annotated as known metabolites.

Partial least squares discriminant analysis was used to find the variable importance in projection (VIP) scores of each annotated metabolite. The fifteen metabolites with the highest VIP scores (Fig. 3 and Fig. A3) showed little contamination from media, as determined by analysis of unused media along with mutant samples. Most of these metabolites were below the limit of detection in unused media, and most of the rest were present at a >30-fold lower abundance in media than in samples, suggesting little or no contamination from residual media (Supplemental data S3). Two metabolites (cytidine and thiamine) were found at similar abundance in media and samples, suggesting these media contamination.

Among the 15 metabolites with high VIP scores, the JCVISYN3A_0728 knockout showed significantly higher abundance of glycerophosphate, oleoyl lysophosphatidic acid, and palmitoylglycerol than other genotypes (Fig. 3 and Fig. A3). We were not able to determine which form of glycerophosphate was increased, although the 3-phosphate is a priori more likely,

being found in the metabolic model as a cardiolipin metabolism intermediate that is synthesized via phosphorylation of imported glycerol by GlpK (JCVISYN3A_0218). To further analyze the knockout metabolomics data further, all four HAD hydrolase knockout phenotypes were separately compared to wild type JCVI-Syn3A (Supplementary data S3). The conclusions are summarized below and further discussed in the Appendix.

The metabolomics data suggests that JCVISYN3A_0066 is the major dNMPase with activity against the deoxymononucleotides dAMP, dGMP, dUMP, dCMP, dTMP and dIMP, and also the ribomononucleotide IMP. Furthermore, as further discussed in the Appendix, the data also suggest the residual presence of pyrimidine nucleoside phosphorylase (PyNP) activity in JCVI-Syn3A after the known MMSYN1_0734 has been removed. The lack of observed nucleotidase activity for JCVISYN3A_0066 in the *in vitro* substrate screen could be due to the absence of relevant effectors. In contrast it seems that JCVISYN3A_0077 is also a dUMP-specific specific dNMPase that plays a minor role *in vivo* compared to JCVISYN3A_0066. The metabolomics data also suggests that JCVISYN3A_0728 is a glycerol 3-phosphate phosphatase. The other activities detected *in vitro*, if relevant *in vivo*, might not be apparent in the metabolomics data if these substrates do not accumulate in cells. No functional role could be proposed for JCVISYN3A_0907.

Comparative genomics uncovers a possible metabolite repair diphosphatase

The YqeK HD family phosphohydrolase is fused to nicotinic acid mononucleotide adenylyltransferase (NadD) in most mycoplasmas and strongly physically clustered with NadD in many other gram-positive organisms (32) (Fig. 4 and Fig. A4A). These genomic associations led us to propose that YqeK repairs mistakes made by NadD. The canonical activity of NadD is to adenylate nicotinate-ribonucleotide (NaMN) using ATP as a donor of the AMP moiety (Fig. 4A). However, use of another NTP or the deoxy-form of ATP would create an erroneous product requiring disposal, most likely by hydrolysis. We therefore expressed JCVISYN3A_0380 and its His230Ala variant in *E. coli* (Fig. A4B). (The His230Ala mutation is predicted to abolish phosphatase activity that would interfere with NadD activity measurement.) *Bacillus subtilis* NadD was used as a benchmark. The JCVISYN3A_0380 His230Ala protein and *B. subtilis* NadD were tested for *in vitro* activity with various nucleoside triphosphates as substrates. The adenylating activity of the JCVISYN3A_0380 His230Ala mutant was quite non-specific and

actually greater against dATP, CTP, or UTP than against the physiological substrate, ATP, whereas *B. subtilis* NadD strongly preferred ATP (Fig. 5A). JCVI-syn3 NadD can therefore readily form deoxy-adenosine, -cytidine, or -uridine analogs of the NAD precursor nicotinate adenine dinucleotide (NaAD), which can presumably be converted to inhibitory analogs of NAD and NADP.

We then tested the JCVI-Syn3 YqeK domain for diphosphatase activity using the NaAD analogs that could be produced by JCVI-syn3A NadD. The YqeK domain had activity towards the cytosine (NaCD) and uracil (NaUD) analogs of NaAD that was at least as high as that against NaAD itself (Fig. 5B), which agrees with the preference of the NadD domain to form these analogs.

We also observed that the YqeK domain had high activity against 8-oxo-GTP, although judging from relative activities with 0.05 mM and 0.5 mM substrate, the K_M is likely higher than for the other substrates tested (Fig. 5B). Consistent with this finding, we showed that the genes encoding the JCVI-syn3A NadD-YqeK fusion can partially complement the *E. coli mutT* high mutation rate phenotype (measured as Rif^R ratios) (Fig. 5C). The partial complementation was also observed when expressing the YqeK domain alone, but not the NadD domain alone. Finally, it was recently shown that YqeKs of gram-positive bacteria belong to a novel diadenosine tetraphosphate (Ap4A) hydrolase family (33). Taken together, these observations suggest that YqeK is a versatile diphosphatase with several functional roles.
Indeed, the available transposon insertion data ((3) and Supplemental data S1) suggested that the NadD domain is essential and the YqeK domain is quasi-essential because a few hits in the YqeK region of the gene were detected in the first Tn round and disappeared after the fourth round. We could not isolate a JCVISYN3A_0380 deletant despite several attempts. We were, however, able to construct a strain carrying the His230Ala mutation that inactivates YqeK diphosphatase activity (Supplemental data S2), and this strain showed no growth defect or obvious metabolite imbalance (Fig. A2).

Metabolomics-driven exploration of damage and repair chemistry in JCVI-Syn3

Thus far, all of our damage and repair cases began with analysis of genes in the JCVI-Syn3A genome and uncovered clear instances of metabolite damage and repair. But are these examples isolated exceptions, or the tip of an iceberg of uncharacterized metabolic chemistry?

To address this question, we adopted a systematic exploratory approach based on the metabolomics data for JCVI-Syn3A cells (see Supplemental Table S3). Because this approach begins with the observed chemical results of potential metabolite damage and is not limited by our current knowledge of gene function, it will certainly find damage mechanisms that our gene-first approach will miss. Still, this approach will also miss any damage mechanisms that fail to be observed through metabolomics, either due to volatility of end products or extremely effective damage mitigation systems.

We focused specifically on a set of 480 metabolites (Supplemental Table S4E) that satisfied two criteria: (1) the mass spectral signal was confidently identified with a defined molecular structure; and (2) the metabolite was at least as abundant in the JCVI-Syn3A cells as in the growth medium. We compared the 480 identified peaks to the 304 metabolites in the JCVI-Syn3A model and the 33,978 compounds in the ModelSEED database (34), resulting in 57 matches to the model and 217 (45%) matches to the database (Supplemental Table S4E). The comparison to the JCVI-Syn3A model reveals two types of discrepancy: (1) 247 metabolites in our model do not appear in our metabolomics data, which is to be expected as many metabolites are too low in concentration or too volatile to be detected in metabolomics; and (2) 423 metabolites that were observed and do not appear in our model, which is more problematic as this implies that there is significant chemistry taking place in this system that our present model cannot explain. The ModelSEED database lookup reveals further discrepancies: (1) 263 observed metabolites do not appear in biochemistry databases, indicating that these is no known biochemical route to any of these compounds that are observed to arise in a biological system; and (2) 160 observed metabolites (217-57) do have known biochemical biosynthesis mechanisms but these mechanisms do not appear in our current JCVI-Syn3 model (3). To predict potential chemical routes to as many of the observed metabolites as possible without limiting our search to known chemistry or straying too far from known JCVI-Syn3A metabolism, we used the PickAxe tool (35). This tool applies generalized reaction rules based on known spontaneous (8) and enzymatic (36, 37) chemical mechanisms to predict potential novel reactions that a given set of metabolites (here, all JCVI-Syn3A metabolites) could undergo. We started with the 304 metabolites present in the JCVI-Syn3A model and applied PickAxe for multiple iterations to allow generation of multistep pathways (see Methods). We used both spontaneous and enzymatic reaction rules in the PickAxe expansion, enabling prediction of pathways with a mixture of both

(as occurs in many damage and repair pathways). The initial PickAxe iterations uncovered an increasing number of compounds generated that matched the observed metabolites, but these hits tapered off after six iterations to just one new compound produced that matched an observed metabolite (blue line in Fig. 6). The number of compounds predicted by PickAxe that matched known biochemistry in the ModelSEED database (green line in Fig. 6) followed a similar trend. We halted the PickAxe expansion at this stage, given its diminishing returns. The final chemical network generated by PickAxe included 33,934 compounds, 61,939 reactions, and matched a total of 182 distinct metabolites (including the original 57 matching the JCVI-Syn3 model) and 1090 ModelSEED compounds (Supplemental data S4C-D).

Next, we used a new flux balance analysis formulation, metabo-FBA, to select a minimal subset of these reactions that connect the functioning JCVI-Syn3A model to as many observed metabolites as possible using mass and energy balanced pathways (see Methods). Because our study is of a minimal genome with relatively few enzymes and specifically focuses on metabolite damage, we favored solutions that involved as many reactions generated by spontaneous reaction rules as possible. This approach produced a predicted flux profile that simultaneously pushed flux through reactions involving compounds that matched 182 observed metabolites (see solution depicted in Fig. 7 and data in Supplemental data S4A and E). This solution included 145 (58%) of the 252 reactions in the JCVI-Syn3 model (purple reactions in Fig. 7), 129 additional ModelSEED reactions (primarily predicted enzymatic reactions; green reactions in Fig. 7), 84 novel enzymatic reactions (blue reactions in Fig. 7), and 74 novel spontaneous reactions (red reactions in Fig. 7) (data in Supplemental data S4A). The fixed image of our flux solution depicted in Fig. 7 is of limited value for permitting a detailed exploration of the fluxes, so we are also including all data files and instructions needed to replicate this view in a fully functioning dynamic Escher map (see Supplemental data S5). Also, the fully expanded version of the JCVI-Syn3A model used to generate this flux solution is provided in SBML and JSON format in Supplemental data S5.

This flux solution is only one of many possible solutions that can explain the observed metabolomics data. While it is unlikely that this solution is completely correct, the true solution must make use of similar chemistry, start with the same initial high-confidence JCVI-Syn3A compounds, and produce the same observed metabolic intermediates, meaning the true solution cannot differ very substantially from our selected one.

The map broadly (Fig. 7), shows clear hotspots of chemical expansion (adenine, cytosine, sugars, pyruvate, amino acids, central carbon trunk reactions, CoA) and regions with little or none (deoxynucleotides, guanine, thymidine, THF, riboflavin, NAD). This is probably explained by the intrinsic reactivities and the concentrations of the associated compounds. Many of the hotspot compounds are high-concentration metabolic starting points (e.g., sugars), end points (e.g., amino acids), or high-flux intermediates (e.g., pyruvate). Their high concentrations make it more likely that these compounds will react chemically and that metabolomics will detect the resulting products.

The many ModelSEED reactions and predicted novel enzymatic reactions proposed by this approach represent previously unannotated but potential promiscuous side activities of existing annotated gene products in JCVI-Syn3A. The metabolomic evidence for the presence of the products of these reactions points strongly to the presence of the reactions themselves. The cluster of ModelSEED reactions expanding from the glucose-6-phosphate (g6p) node of the JCVI-Syn3A model is a good example (see Fig. 7). These reactions are phosphorylations and hydrolyses that interconvert many diverse sugars and polysaccharides, all of which are supported by our metabolomics data. As the model only contains reactions for glucose as a representative sugar, it probably understates the extent of such reactions.

Also of note, is how many of the pathways predicted in JCVI-Syn3A by our metabo-FBA method involve a mixture of database reactions, predicted spontaneous reactions, and novel enzymatic reactions (30/50 total pathways). Any analysis based on just one or two of these three reaction sources would explain a far smaller number of observed metabolites due to gaps and dead-ends in the predicted pathways. This, using all three reaction data sources provides a much fuller understanding of metabolism.

Another notable point is that much of the new predicted chemistry surrounds amino acids. Many of the observed metabolomics peaks correspond to amino acid derivatives such asdipeptides and acetylated amino acids (see Fig. 7). The dipeptides serve primarily as nutrients for JCVI-syn3, which contains the peptidases needed to degrade these compounds (a large number of the ModelSEED reactions added by our metabo-FBA approach relate to dipeptide transport and degradation). The acetylated amino acids are different in that only 7 out of 10 of these compounds were found in biochemistry databases, and the databases lacked spontaneous acetylation reactions to produce these compounds. Yet, metabolomics evidence supported the

presence of all 10 in the JCVI-Syn3A strain. The metabo-FBA approach added 10 predicted spontaneous acetylation reactions, using acetyl-phosphate as a donor, based on PickAxe predictions. This demonstrates how readily acetylation occurs in these systems, either by spontaneous action or by promiscuous enzyme activity, and it highlights the particular vulnerability of amino acids to this acetylation.

These results also support previous hypotheses about the main metabolic network of JCVI-Syn3A (3) with regard to acetyl phosphate and the enzymes producing/consuming it. The in vivo essentiality of phosphate acetyltransferase (JCVISYN3_0229) and acetate kinase (JCVISYN3_0230) was previously puzzling, given that the upstream genes in the pathway, the subunits of pyruvate dehydrogenase (JCVISYN3_0227/8), were found to be non-essential *in vivo*. It had been hypothesized that the two former enzymes were essential because buildups of acetyl-CoA or acetyl phosphate needed to be prevented, both being known protein acetylation agents (38). The extensive and diverse acetylation damage we found evidenced in our metabolomics data would seem to further support this hypothesis.

Relatedly, our results support a role for acetyl phosphate in the acetylation of proteins as well as free amino acids because some of the identified amino acids had side chain acetylations. The results also support the hypothesized essential role of acetate kinase as a means of preventing acetyl phosphate accumulation.

These analyses also expose insights into the relative importance of our various proposed mechanisms for spontaneous chemistry, based on which mechanisms are most likely to give rise to metabolic products found in our metabolomics data (see larger discussion in Appendix and Fig. A5). Of course, not all chemically impactful metabolites are readily observed in metabolomics data due to instability or volatility. Methyl-glyoxal is a good example of an important metabolite that arises from and participates in spontaneous damage reactions but could not be observed (Fig. 7). While methylglyoxal was not among the observed metabolites due to small size and volatility, metabo-FBA added reactions involving this compound because it leads to numerous downstream potential damage and repair reactions. A more detailed discussion of methylglyoxal follows.

Possible ways for JCVI-Syn3A to cope with methylglyoxal stress

Methylglyoxal is necessarily formed from the triose phosphates in JCVI-Syn3A central metabolism (39) but the classical glyoxalase system comprising the glutathione-dependent GloA and GloB enzymes (40) is absent. Likewise, the JCVI-Syn3A genome does not encode enzymes with minor methylglyoxal-detoxifying activities, such as aldose reductases and keto-aldehyde reductases (41–43). The only candidate enzyme that we identified as potentially able to counter methylglyoxal-induced damage is JCVISYN3A_0400, which encodes a homolog of DJ-1. The DJ-1 superfamily has several functionally distinct clades, of which four are found in *E. coli* (encoded by *hch*A, *yaj*L, *yhb*O and *elb*B). Phylogenetic analysis places JCVISYN3A_0400 in the YajL/DJ-1 clade (Fig. A6).

The members of the DJ-1 superfamily that have been functionally characterized participate in stress response and detoxification (44). Some are thought to be deglycases (45), glyoxalases (46), or aldehyde-adduct hydrolases (47). Previous studies showed variability in the phenotypes reported for the *E. coli hch*A, *yaj*L, *yhb*O deletion mutants as the sensitivity of the *yaj*L reported by the Richarme group (48) was not reproduced in independent studies (46). We also failed to reproduce the reported glyoxal or methylglyoxal sensitivities of the single deletion *yajL* mutant, but did observe a defect both in its growth rate and yield of the the Δ*yaj*L/Δ*hch*A *E. coli* K-12 BW25113 strain (Fig. 8A and Fig. S7A). Expression of the *E. coli yajL* or JCVISYN3A_0400 genes *in trans* complemented this growth phenotype (Fig. 8A and Fig. A7A) suggesting JCVISYN3A_0400 is indeed in the same DJ-1 subgroup as YajL.

To test the hypothesis that JCVISYN3A_0400 participates in methylglyoxal detoxification, we measured the glyoxalase activity of the recombinant protein. JCVISYN3A_0400 possess a low but measurable methylglyoxalase activity (k_{cat} = 0.025 ± 0.002 sec^{-1}, K_M = 1.23 ± 0.30 mM), lower than obtained for the positive control protein human DJ-1 (k_{cat} = 0.126 ± 0.004 sec^{-1}, K_M = 0.34 ± 0.04 mM) but higher than *E. coli* YajL (k_{cat} = 0.004 ± 0.0001 sec^{-1}, K_M = 0.095 ± 0.018 mM) (Fig. 8B). The low k_{cat} for YajL is consistent with a prior report that did not detect glyoxalase activity using methylglyoxal as a substrate (46). The ~20 M^{-1} sec^{-1} k_{cat}/K_M value for JCVISYN3A_0400 is five to six orders of magnitude lower than that of glyoxalase I, the canonical glutathione-dependent glyoxalase (49). Even compared to other DJ-1 superfamily glyoxalases, JCVISYN3A_0400 is a poor enzyme. The lactate oxidase-coupled assay used here is specific to L-lactate, which should detect all the lactate produced by JCVISYN3A_0400, as a prior study indicated that DJ-1 clade enzymes produce only the L

enantiomer (50), although we did not test the enantiopurity of the lactate produced by JCVISYN3A_0400 in this study.

Because DJ-1 superfamily members have been reported to be generalist deglycases (51), we tested the deglycase activity of JCVISYN3A_0400 against the methylglyoxal-CoA hemithioacetal (Fig. A7B). CoA was used as the thiol because the absence of glutathione biosynthetic enzymes in JCVI-Syn3A suggests that CoA is its main small molecule thiol (see above). JCVISYN3A_0400 had no detectable deglycase activity against methylglyoxal-CoA hemithioacetal (Fig. A7B), while human DJ-1 had a low activity (k_{cat} = 0.0068 ± 0.0007 sec^{-1}, K_M = 0.144 ± 0.064 mM) against the same substrate (52). JCVISYN3A_0400 therefore seems unlikely to efficiently detoxify methylglyoxal via either glyoxalase or deglycase pathways. It is possible that JCVISYN3A_0400 and other DJ-1-type glutathione-independent methylglyoxalases have some unidentified positive effector *in vivo* that enhances their activity, and the glyoxalase activity of human DJ-1 is highly sensitive to buffer conditions (53). In summary, while results suggest that JCVISYN3A_0400 and YajL are iso-functional, they do not appear to make a large contribution to methylglyoxal detoxification.

The recent observation that human DJ-1, *E. coli* Yajl, and *S. pombe* DJ-1 can reduce the levels of modifications derived from 1,3 bisphosphoglycerate suggests an alternative hypothesis for the function of JCVISYN3A_0400 and other close DJ-1 homologs (54). It is possible that these proteins share an evolutionarily conserved function in detoxifying an electrophilic cyclic 1,3 phosphoglycerate intermediate that is spontaneously formed by intramolecular cyclization of 1,3 bisphosphoglycerate (54). This metabolite should be formed in all organisms that use glycolysis and thus provides a possible explanation for why the minimal Mycoplasma JCVI-Syn3A would need to preserve this pathway.

Conclusion

Metabolite damage arising from side-reactions of enzymes and spontaneous chemistry has often been ignored or seen as a minor metabolic inconvenience – even a trivial sideshow – that does not warrant investment in enzymes to prevent or repair it (6). Biochemical, genetic, and engineering evidence accumulating over the past decade have started to change this view (6, 8, 13, 15, 55, 56). The biochemical and genetic results presented here constitute persuasive additional evidence by demonstrating that stripping a genome down to its barest essentials leaves

metabolite damage-control systems in place. Furthermore, our metabolomic and cheminformatic results point to the existence of a network of metabolite damage and damage-control reactions that extends far beyond the corners of it characterized so far. In sum, there can be little room left to doubt that damage itself and the systems that counter it are mainstream metabolic processes.

Methods

Bioinformatics

The BLAST tools (57) and CDD resources at NCBI (http://www.ncbi.nlm.nih.gov/) (58) were routinely used. Sequences were aligned using Clustal Omega (59) or Multialin (60). Phylogenetic distribution was analyzed in the SEED database (61). Results are available in the "YqeK" subsystem on the PubSEED server (http://pubseed.theseed.org//SubsysEditor.cgi?page=ShowSpreadsheet&subsystem=NadD-YqeK_fusion_display). Physical clustering was analyzed with the SEED subsystem coloring tool or the SeedViewer Compare Regions tool (61) and the clustering figure was generated with GeneGraphics (62). Phylogenetic trees were constructed with Mega 6 (63). Student's t-test calculations were performed using the VassarStats web-tools (http://vassarstats.net).

Prediction of novel potential chemistry using PickAxe

Expanded chemistry was generated using the PickAxe app in KBase, as shown in this narrative: https://narrative.kbase.us/narrative/29280. This app uses the open source RDKit package to apply sets of SMARTS-based chemical reaction rules, derived from previously published chemical damage (8) and enzyme promiscuity (35) studies, to an input set of compounds to produce all possible reactions and products that might arise from that chemistry. This analysis can be run iteratively through repeated application of the reaction rules to all new products that arise from previous generations. We applied the PickAxe approach for six iterations, retaining all compounds that matched the JCVI-Syn3A model, the ModelSEED database (34), or an observed metabolite.

Metabo-flux balance analysis to predict minimal reactions to reach observed metabolites

In metabo-flux balance analysis (metabo-FBA), constraints are added to the standard FBA formulation to force flux through one or more reactions involving an observed metabolite. In this

formulation, a variable is added for each observed peak (p_i) and a variable is added for each metabolite that has been mapped to the peak (because peaks lack stereochemistry, they may be mapped to multiple possible stereoisomers). Next, a constraint is added stating that a peak cannot be active unless one or more of its associated metabolites is active (where $\lambda_{i,j}$ is a mapping variable equal to 1 if metabolite *j* is mapped to peak *i* and zero otherwise):

$$p_i \leq \sum_{j}^{Compounds} \lambda_{i,j} m_j$$

A constraint is also added stating that no metabolite can be active unless at least one reaction in which the metabolite is involved is carrying flux (where $\gamma_{j,k}$ is a mapping variable equal to 1 if metabolite *j* is involved in reaction *k* and zero otherwise):

$$m_j \leq \sum_{k}^{Reactions} 100\gamma_{j,k} v_k$$

To maximize active metabolites, the objective of the problem is then set to maximize the sum of all p_i. While p_i and m_j can be specified as binary variables, it works equally well and is less computationally expensive to use continuous variables bounded between 0 and 0.1. To avoid the trivial solution of activating metabolites by pushing flux through both directions of reversible reactions or around mass balanced flux loops, it is essential to also employ thermodynamics constraints in some form in this formulation (64).

Media, strains, and genetic manipulations

All strains, plasmids and oligonucleotides used in this study are listed in Table A4 and Table A5. Bacterial growth media were solidified with 15 g/l agar (BD Diagnostics Systems) for the preparation of plates. *E. coli* were routinely grown on LB medium (BD Diagnostics Systems) at 37 °C unless otherwise stated. Transformations were performed following standard procedures (62). IPTG (100 μM), Ampicillin (Amp, 100 μg/ml), Kanamycin (Km, 50 μg/ml), l-Arabinose (Ara, 0.02–0.2%), Chloramphenicol (Cm, 25 μg/ml) and Rifampicin (Rif, 25 μg/ml) were used when appropriate. Bacterial M9 minimal medium (65), 0.4% (w/v) glucose was used either with NH_4Cl (20 mM) or glycine (50 mM) as the nitrogen source. P1 transduction was performed following the classical methods (66). The Kan^R marker was eliminated from the BW2113 Δ*yaj*L::Kan^R strain by the procedure described by Cherepanov and Wackernagel (67).

Transductants from BW2113 Δ*hch*A::Kan^R to BW2113Δ*yaj*L were checked by PCR for transduction of the Δ*hch*A::Kan^R allele into the recipient strains using primer pairs [DH492/493 (ext); DH494/495 (int) and DH480/481 (ext); DH482/483 (int)] respectively. Plasmid constructions for expression JCVI-syn3A genes in *E. coli* are described in the supplemental methods.

JCVI-syn3A is a near minimal bacterial cell first reported by Breuer *et al.* (3)that contains a subset of the genes in *Mycoplasma mycoides* subspecies *capri* strain GM12. Mycoplasmas were grown in SP4 broth (68) that contains 17% KnockOut Serum Replacement™ instead of 17% fetal bovine serum and is referred to as SP4-KO as described in the supplemental Methods. Construction of gene knockout mutants in JCVI-Syn3A was a multistep process, and two different protocols were used. These protocols are described in detail in the **Supplemental data S2** file.

Mutation frequency assays for *E. coli* derivatives

Overnight cultures in LB with added antibiotics and arabinose (0.02%) were diluted 100-fold in the same conditions and grown for another 24 h before dilutions were plated on LB and LB rifampicin (25 μg/ml) to calculate a mutation ratio (Number of colonies on Rif x dilution factor) / (Number of colonies on LB x dilution factor).

Protein expression and purification and enzyme assays

All characterized JCVI-syn3A encoded proteins were expressed as His-tagged variants in *E. coli* and purified using Ni^{2+}-NTA columns as described in Supplemental Methods. In vitro activity assays for CoA disulfide reductase, for phosphatase with a range of substrates, NadD, glyoxalase, and deglycase are described in detail in Supplemental Methods.

The appendix and supplemental data have been deposited in the Figshare data depository with the DOI: 10.6084/m9.figshare.20020574.

Acknowledgements. This work was funded by the National Science Foundation (Grants MCB-1611846 to OF, MCB-1611952 to CH and MCB-1611711 to ADH and V dC-L, MCB-1840301, MCB-1840320 and MCB-1818344 subcontracts to J.I.G.) and by the J. Craig Venter Institute.

Table 1. Members of the HAD family of unknown function encoded by JCVI-Syn3

Gene	Family	Essential	Best 3 substrates Activity *in vitro**	Physical clustering	*M. florum* ortholog locus tag and essentiality**
JCVISYN3A_0066	Cof subfamily of IIB subfamily of HAD superfamily	no	pNPP, FMN, CoA	Between 5S rRNA gene and thioredoxin	Mfl169 (NE)
JCVISYN3A_0077	Cof-like hydrolase, HAD superfamily	no	Fru-1P, Ery-4P	Between *tsaD* and *aspS*	Mfl614 (E)
JCVISYN3A_0710	Cof subfamily of IIB subfamily of HAD superfamily	yes	Could not clone	Between tRNA genes and predicted phosphonate transporter genes	Mfl513 (E)
JCVISYN3A_0728	HAD superfamily hydrolase subfamily IIB, protein	no	GMP XMP 2-deoxy-glucose-6P	Between glycolysis genes	Mfl503 (E)
JCVISYN3A_0907	Cof-like hydrolase, HAD superfamily	no	N-acetyl-D-glucosamine-6P Fructose-1P N-acetyl-D-glucosamine-1P	Between YidC and choline kinase-like	Mfl680 (NE)

*Abbreviations in Table S1; ** (E), essential; (NE)=non-essential in *M. florum*

References

1. Schwille P. 2011. Bottom-up synthetic biology: engineering in a tinkerer's world. Science (80-) 333:1252–1254.
2. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi Z-Q, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC. 2016. Design and synthesis of a minimal bacterial genome. Science 351:aad6253.
3. Breuer M, Earnest TM, Merryman C, Wise KS, Sun L, Lynott MR, Hutchison CA, Smith

HO, Lapek JD, Gonzalez DJ, de Crécy-Lagard V, Haas D, Hanson AD, Labhsetwar P, Glass JI, Luthey-Schulten Z. 2019. Essential metabolism for a minimal cell. Elife 8:e36842.

4. Danchin A, Fang G. 2016. Unknown unknowns: essential genes in quest for function. Microb Biotechnol 9:530–540.

5. Antczak M, Michaelis M, Wass MN. 2019. Environmental conditions shape the nature of a minimal bacterial genome. Nat Commun 10:3100.

6. Linster CL, Van Schaftingen E, Hanson AD. 2013. Metabolite damage and its repair or pre-emption. Nat Chem Biol 9:72–80.

7. Galperin MY, Moroz O V, Wilson KS, Murzin AG. 2006. House cleaning, a part of good housekeeping. Mol Microbiol 59:5–19.

8. Lerma-Ortiz C, Jeffryes JG, Cooper AJL, Niehaus TD, Thamm AMK, Frelin O, Aunins T, Fiehn O, de Crécy-Lagard V, Henry CS, Hanson AD. 2016. "Nothing of chemistry disappears in biology": the Top 30 damage-prone endogenous metabolites. Biochem Soc Trans 44:961–71.

9. de Crécy-Lagard V, Haas D, Hanson AD. 2018. Newly-discovered enzymes that function in metabolite damage-control. Curr Opin Chem Biol 47:101–108.

10. Niehaus TD, Richardson LG, Gidda SK, ElBadawi-Sidhu M, Meissen JK, Mullen RT, Fiehn O, Hanson AD. 2014. Plants utilize a highly conserved system for repair of NADH and NADPH hydrates. Plant Physiol 165:52–61.

11. Bommer GT, Van Schaftingen E, Veiga-da-Cunha M. 2020. Metabolite Repair Enzymes Control Metabolic Damage in Glycolysis. Trends Biochem Sci 45:228–243.

12. Becker-Kettern J, Paczia N, Conrotte JF, Zhu C, Fiehn O, Jung PP, Steinmetz LM, Linster CL. 2018. NAD(P)HX repair deficiency causes central metabolic perturbations in yeast and human cells. FEBS J 285:3376–3401.

13. Van Schaftingen E, Rzem R, Marbaix A, Collard F, Veiga-Da-Cunha M, Linster CL. 2013. Metabolite proofreading, a neglected aspect of intermediary metabolism. J Inherit Metab Dis 36:427–434.

14. Veiga-da-Cunha M, Van Schaftingen E, Bommer GT. 2020. Inborn errors of metabolite repair. J Inherit Metab Dis 43:14–24.

15. Sun J, Jeffryes JG, Henry CS, Bruner SD, Hanson AD. 2017. Metabolite damage and

repair in metabolic engineering design. Metab Eng 44:150–159.
16. Hanson AD, Henry CS, Fiehn O, de Crécy-Lagard V. 2016. Metabolite damage and metabolite damage control in plants. Annu Rev Plant Biol 67:131–52.
17. Stover P, Schirch V. 1990. Serine hydroxymethyltransferase catalyzes the hydrolysis of 5,10-methenyltetrahydrofolate to 5-formyltetrahydrofolate. J Biol Chem 265:14227–14233.
18. Stover P, Schirch V. 1993. The metabolic role of leucovorin. Trends Biochem Sci 18:102–106.
19. Jeanguenin L, Lara-Nùñez A, Pribat A, Hamner Mageroy M, Gregory 3rd JF, Rice KC, de Crécy-Lagard V, Hanson AD. 2010. Moonlighting glutamate formiminotransferases can functionally replace 5-formyltetrahydrofolate cycloligase. J Biol Chem 285:41557–66.
20. Poole LB. 2015. The basics of thiols and cysteines in redox biology and chemistry. Free Radic Biol Med 80:148–57.
21. Berglund O, Holmgren A. 1975. Thioredoxin reductase-mediated hydrogen transfer from *Escherichia coli* thioredoxin-(SH)2 to phage T4 thioredoxin-S2. J Biol Chem 250:2778–2782.
22. Thelander L. 1967. Thioredoxin Reductase: Characterization of a homogeneous preparation froa Escherichia coli B. J Biol Chem 242:852–859.
23. Holmgren A. 1985. Thioredoxin. Annu Rev Biochem 54:237–271.
24. Ben-Menachem G, Himmelreich R, Herrmann R, Aharonowitz Y, Rottem S. 1997. The thioredoxin reductase system of mycoplasmas. Microbiology 143:1933–1940.
25. Thelander L, Reichard P. 1979. Reduction of ribonucleotides. Annu Rev Biochem 48:133–158.
26. DelCardayré SB, Stock KP, Newton GL, Fahey RC, Davies JE. 1998. Coenzyme A disulfide reductase, the primary low molecular weight disulfide reductase from *Staphylococcus aureus* : purification and characterization of the native enzyme. J Biol Chem 273:5744–5751.
27. Benyoucef M, Rigaud J-L, Leblanc G. 1981. The electrochemical proton gradient in Mycoplasma cells. Eur J Biochem 113:491–498.
28. Seifried A, Schultz J, Gohla A. 2013. Human HAD phosphatases: structure, mechanism, and roles in health and disease. FEBS J 280:549–571.

29. Lachance J-C, Matteau D, Brodeur J, Lloyd CJ, Mih N, King ZA, Knight TF, Feist AM, Monk JM, Palsson BO, Jacques P-É, Rodrigue S. 2021. Genome-scale metabolic modeling reveals key features of a minimal gene set. Mol Syst Biol 17:e10099–e10099.
30. Papenfort K, Sun Y, Miyakoshi M, Vanderpool CK, Vogel J. 2013. Small RNA-mediated activation of sugar phosphatase mRNA regulates glucose homeostasis. Cell 153:426–437.
31. Huang L, Khusnutdinova A, Nocek B, Brown G, Xu X, Cui H, Petit P, Flick R, Zallot R, Balmant K, Ziemak MJ, Shanklin J, De Crécy-Lagard V, Fiehn O, Gregory JF, Joachimiak A, Savchenko A, Yakunin AF, Hanson AD. 2016. A family of metal-dependent phosphatases implicated in metabolite damage-control. Nat Chem Biol 12.
32. Aravind L, Koonin E V. 1998. The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 23:469–72.
33. Minazzato G, Gasparrini M, Amici A, Cianci M, Mazzola F, Orsomando G, Sorci L, Raffaelli N. 2020. Functional characterization of COG1713 (YqeK) as a novel diadenosine tetraphosphate hydrolase family. J Bacteriol 202:e00053-20.
34. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber ME, Best AA, DeJongh M, Kimbrel JA, D'haeseleer P, McCorkle SR, Bolton JR, Pearson E, Canon S, Wood-Charlson EM, Cottingham RW, Arkin AP, Henry CS. 2020. The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Res 49:D1555.
35. Jeffryes JG, Colastani RL, Elbadawi-Sidhu M, Kind T, Niehaus TD, Broadbelt LJ, Hanson AD, Fiehn O, Tyo KEJ, Henry CS. 2015. MINEs: Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. J Cheminform 7:44.
36. Hatzimanikatis V, Li C, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. 2005. Exploring the diversity of complex metabolic networks. Bioinformatics 21:1603–9.
37. Henry CS, Broadbelt LJ, Hatzimanikatis V. 2010. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. Biotechnol Bioeng 106:462–73.
38. Shi L, Tu BP. 2015. Acetyl-CoA and the regulation of metabolism: mechanisms and consequences. Curr Opin Cell Biol 33:125–131.

627 39. Phillips SA, Thornalley PJ. 1993. The formation of methylglyoxal from triose phosphates.
628 Eur J Biochem 212:101–105.
629 40. Sukdeo N, Honek JF. 2008. Microbial glyoxalase enzymes: Metalloenzymes controlling
630 cellular levels of methylglyoxal. Drug Metabol Drug Interact 23:29–50.
631 41. Inoue Y, Kimura A. 1995. Methylglyoxal and regulation of its metabolism in
632 microorganisms. Adv Microb Physiol 37:177-227.
633 42. Misra K, Banerjee AB, Ray S, Ray M. 1996. Reduction of methylglyoxal in *Escherichia*
634 *coli* K12 by an aldehyde reductase and alcohol dehydrogenase. Mol Cell Biochem
635 156:117–24.
636 43. Jagt DLV, Robinson B, Taylor KK, Hunsaker LA. 1992. Reduction of trioses by NADPH-
637 dependent aldo-keto reductases. Aldose reductase, methylglyoxal, and diabetic
638 complications. J Biol Chem 267:4364–9.
639 44. Smith N, Wilson MA. 2017. Structural biology of the DJ-1 superfamily. Adv Exp Med
640 Biol 1037:5–24.
641 45. Richarme G, Liu C, Mihoub M, Abdallah J, Leger T, Joly N, Liebart J-C, Jurkunas U V,
642 Nadal M, Bouloc P, Dairou J, Lamouri A. 2017. Guanine glycation repair by DJ-1/Park7
643 and its bacterial homologs. Science (80-) 357:208–211.
644 46. Lee C, Lee J, Lee JY, Park C. 2015. Characterization of the *Escherichia coli* Yajl, YhbO
645 and ElbB glyoxalases. FEMS Microbiol Lett 363:fnv239.
646 47. Matsuda N, Kimura M, Queliconi BB, Kojima W, Mishima M, Takagi K, Koyano F,
647 Yamano K, Mizushima T, Ito Y, Tanaka K. 2017. Parkinson's disease-related DJ-1
648 functions in thiol quality control against aldehyde attack in vitro. Sci Rep 7:12816.
649 48. Abdallah J, Mihoub M, Gautier V, Richarme G. 2016. The DJ-1 superfamily members
650 YhbO and YajL from *Escherichia coli*repair proteins from glycation by methylglyoxal
651 and glyoxal. Biochem Biophys Res Commun 470:282–286.
652 49. Clugston SL, Barnard JFJ, Kinach R, Miedema D, Ruman R, Daub E, Honek JF. 1998.
653 Overproduction and characterization of a dimeric non-zinc glyoxalase I from *Escherichia*
654 *coli*: Evidence for optimal activation by nickel ions. Biochemistry 37:8754–63.
655 50. Choi D, Kim J, Ha S, Kwon K, Kim EH, Lee HY, Ryu KS, Park C. 2014. Stereospecific
656 mechanism of DJ-1 glyoxalases inferred from their hemithioacetal-containing crystal
657 structures. FEBS J 281:5447–62.

51. Richarme G, Abdallah J, Mathas N, Gautier V, Dairou J. 2018. Further characterization of the Maillard deglycase DJ-1 and its prokaryotic homologs, deglycase 1/Hsp31, deglycase 2/YhbO, and deglycase 3/YajL. Biochem Biophys Res Commun 503:703–709.
52. Mazza MC, Shuck S, Lin J, Moxley MA, Termini J, Cookson MR, Wilson MA. 2022. DJ-1 glyoxalase activity makes a modest contribution to cellular defense against methylglyoxal damage in neurons. bioRxiv 2022.02.18.481064.
53. Mulikova T, Bekkhozhin Z, Abdirassil A, Utepbergenov D. 2021. A continuous spectrophotometric assay for glutathione-independent glyoxalases. Anal Biochem 630:114317.
54. Heremans IP, Caligiore F, Gerin I, Bury M, Lutz M, Graff J, Stroobant V, Vertommen D, Teleman AA, Van Schaftingen E, Bommer GT. 2022. Parkinson's disease protein PARK7 prevents metabolite and protein damage caused by a glycolytic metabolite. Proc Natl Acad Sci U S A 119:e2111338119.
55. Van Schaftingen E, Veiga-da-Cunha M, Linster CL. 2015. Enzyme complexity in intermediary metabolism. J Inherit Metab Dis 38:721–727.
56. Golubev A, Hanson AD, Gladyshev VN. 2017. Non-enzymatic molecular damage as a prototypic driver of aging. J Biol Chem 292:6029–6038.
57. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402.
58. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8–D13.
59. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, Lopez R. 2015. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res2015/04/08. 43:W580-4.
60. Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids 16:10881–10890.
61. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res=. 42:D206-14.

62. Harrison KJ, de Crécy-Lagard V, Zallot R. 2018. Gene Graphics: A genomic neighborhood data visualization web application. Bioinformatics 34:1406–1408.
63. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol2013/10/18. 30:2725–2729.
64. Henry CS, Broadbelt LJ, Hatzimanikatis V. 2007. Thermodynamics-based metabolic flux analysis. Biophys J 92:1792–1805.
65. Sambrook JE, Fritsch EF, Maniatis T, Fritsch EF, Sambrook JE, Fritsch EF, Maniatis T, Fritsch F, Maniatis T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
66. Miller JH. 1972. Experiments in Molecular Genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
67. Cherepanov PP, Wackernagel W. 1995. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158:9–14.
68. Tully JG, Rose DL, Whitcomb RF, Wenzel RP. 1979. Enhanced isolation of *Mycoplasma pneumoniae* from throat washings with a newly modified culture medium. J Infect Dis 139:478–482.

Figure legends

Figure 1. 5-FCL activity is encoded by JCVI_0400. **(A)** Enzymatic source and repair of 5-CHO-THF. **(B)** Growth phenotype of a WT *E. coli* BW25113 , Δ*ygf*A mutant and, Δ*ygf*A mutant expressing JCVI_0443 gene on M9 minimal medium (0.4% glucose) with (**1)** 20 mM NH_4Cl or (**2)** 50 mM glycine as sole nitrogen source. Plates were incubated for 3 days at 37°C.

Figure 2. Predicted and validated redox buffering systems in JCVI-Syn3. (A) Candidates for H_2O_2 detoxification systems of JCVI3, experimentally validated are in solid arrows, only the number of the locus tags are given, P is for protein, R is for small molecule. The predicted source of reductant is NADPH (**B)** CoADR Michaelis-Menten saturation curve for the determination of the Km and kcat for CoAD consumption. **(C)** CoADR is specific towards oxidized CoA with no activity towards other tested disulfides

Figure 3. Heatmap including 15 metabolites from JCVI-syn3A mutant metabolomic analysis with highest VIP scores. Samples and genotypes are represented in columns. High intensity measurements as compared to average intensity are red/yellow, and low intensity measurements are represented by green/blue

Figure 4. Predicted Hydrolase of unknown function is clustered or fused to NadD in many Firmicutes (A) Predicted $NADP^+$ synthesis pathway in JCVI-Syn3. **(B)** Physical clustering and fusions of *nadD* and *ykeK* homologs in several gram-positive Bacteria. The RefSeq identifiers for the *yqeK* genes used in descending order are: NP_975428.1, NP_390441.1, NP_372117.1, NP_816490.1, YP_140036.1. **(C)** Docked model of 2-deoxy-NaAD bound to the *C. acetobutylicum* YqeK (pdb code: 3CCG). The protein is shown in ribbon format (grey) with side chains as lines, two iron atoms are shown as spheres bound to the diphosphate of dNaAD. Tyrosine 82 (green) is modeled as two conformations in the crystal structure and forms a close interaction with the 2'carbon of dNaAD.

Figure 5. Biochemical analysis of the NadD and YqeK activities (A) Relative reaction rates of *Bacillus subtilis* and JCVI syn3.0 NadD enzymes with NaMN and various nucleotides, calculated as percentage of the canonical reaction with ATP for each NadD enzyme. Enzymes

were incubated with 2 mM NTP, 0.5 mM NaMN, 4 mM $MgCl_2$ and 5 u/ml yeast inorganic pyrophosphatase for 5 min at 37° C. H230A has the conserved H in the active site of the YqeK domain mutated to ablate the HD activity and cleavage of nucleotides. **(B)** Activity of the expressed JCVI syn3.0 YqeK domain with different substrates. YqeK (0.2 μg) was incubated with 0.5 or 0.05 mM substrates, 1 mg/ml BSA and 2.0 mM $MgCl_2$ for 20 min at at 37° C. Black bars are data for 0.5 mM substrates, white bars are data for 0.05 mM substrates. **(C)** Mutation ratio on LB rifampicin for strain Δ*mut*T with empty vector (pBAD24), Δ*mut*T with E. coli *mut*T in trans, Δ*mut*T with either the *nad*D-*yqe*K fusion gene JCVI_0380, or the *nad*D or *yqe*K domains alone. *** indicates a P-value <0.001 with experiments performed with four biological replicates and four technical replicates**.**

Figure 6. Number of predicted potential metabolites arising from promiscuous enzymatic reactions and spontaneous/damage chemistry operating on known compounds in JCVI-syn3 metabolism. Total predicted metabolites are shown, as well as the number of metabolites matching observed peaks (blue line) or ModelSEED compounds (green line). The x-axis indicates the number of reactions steps explored outward from the known JCVI-syn3 metabolism, while the y-axis shows the number of new metabolites predicted with each new reaction step.

Figure 7. Map of predicted extensions to the JCVI-syn3 model to push flux through as many observed peaks as possible. Reactions and metabolites are color coded as shown in the figure's inset. Model reactions with no flux (black); and with flux (magenta). Predicted and active reactions that are in the database (green); or that are novel and spontaneous (red); or that are novel enzymatic ones (blue). All active predicted spontaneous reactions and nearly all active model reactions are shown on the map; some ModelSEED and predicted enzymatic reactions are excluded. The color code for metabolites is as follows: absent in the mass spec analysis (white); observed metabolites that are also in the model (yellow); in the database (ochre); or novel in themselves or in the way they are produced (brown). Most enzymatic reactions are identified by their EC numbers. Some reactants' names have been omitted since they don't give relevant information. Common abbreviations have been used for the name labels. The map has been divided by panels shown on the figure's background. These panels are labeled according to the

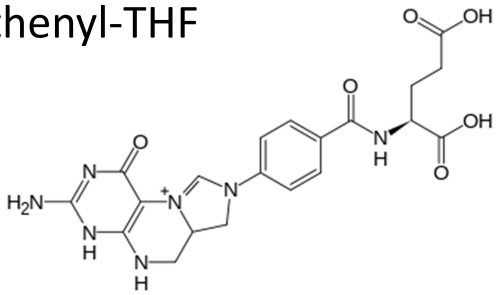
major pathway they display. The complete metabolic map in interactive format (Escher map) is given in the supplemental data S5 material.

Figure 8. Characterization of JCVI_0400. (A) Growth of WT, Δ*yaj*L, Δ*yaj*L Δ*hch*A, Δ*yaj*L Δ*hch*A with *hchA* in trans and Δ*yaj*L Δ*hch*A with JCVI_0400 in trans. pUC19 was used as empty vector. Each strain was tested in 5 replicates Plates were incubated 2 days at 37°C in LB with agitation in a Bioscreen C device. **(B)** Methylglyoxalase activity of JCVIsyn3A_0400 compared to human DJ-1 (DJ1) and *E. coli* YajL. Conversion of methylglyoxal to L-lactate was measured in a coupled assay with L-lactate oxidase and Amplex red. Data were measured in triplicate with error bars shown (sometimes smaller than the symbol) and fitted using the Michaelis-Menten model. JCVIsyn3_0400 is a weak methylglyoxalase.

Figure 1

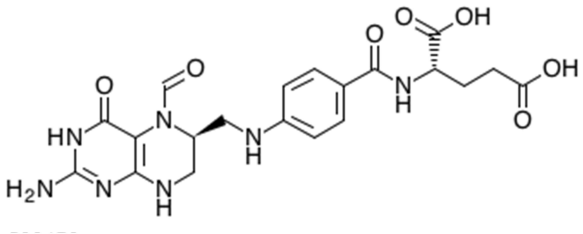
A

5,10-methenyl-THF



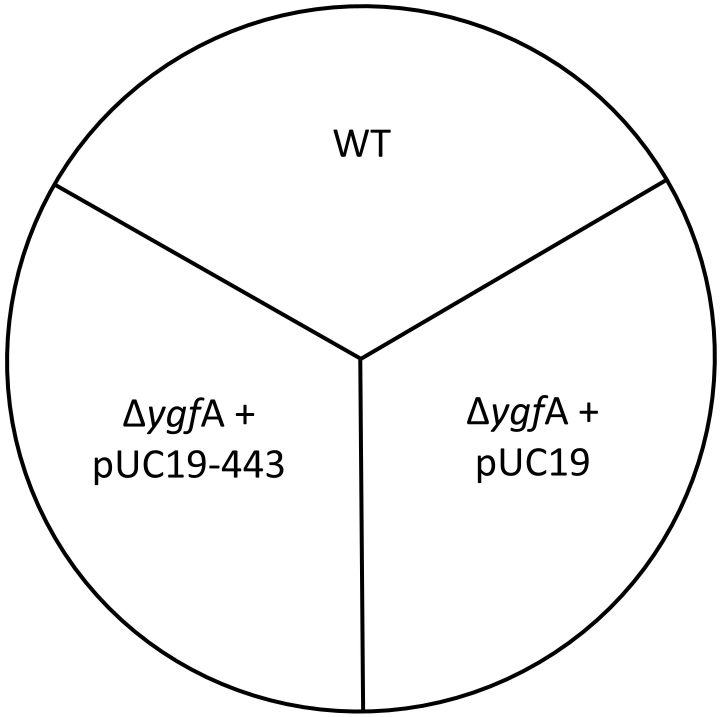
GlyA (SHMT)
damage

YgfA (5-FCL)
repair



5-Formyltetrahydrofolate

B



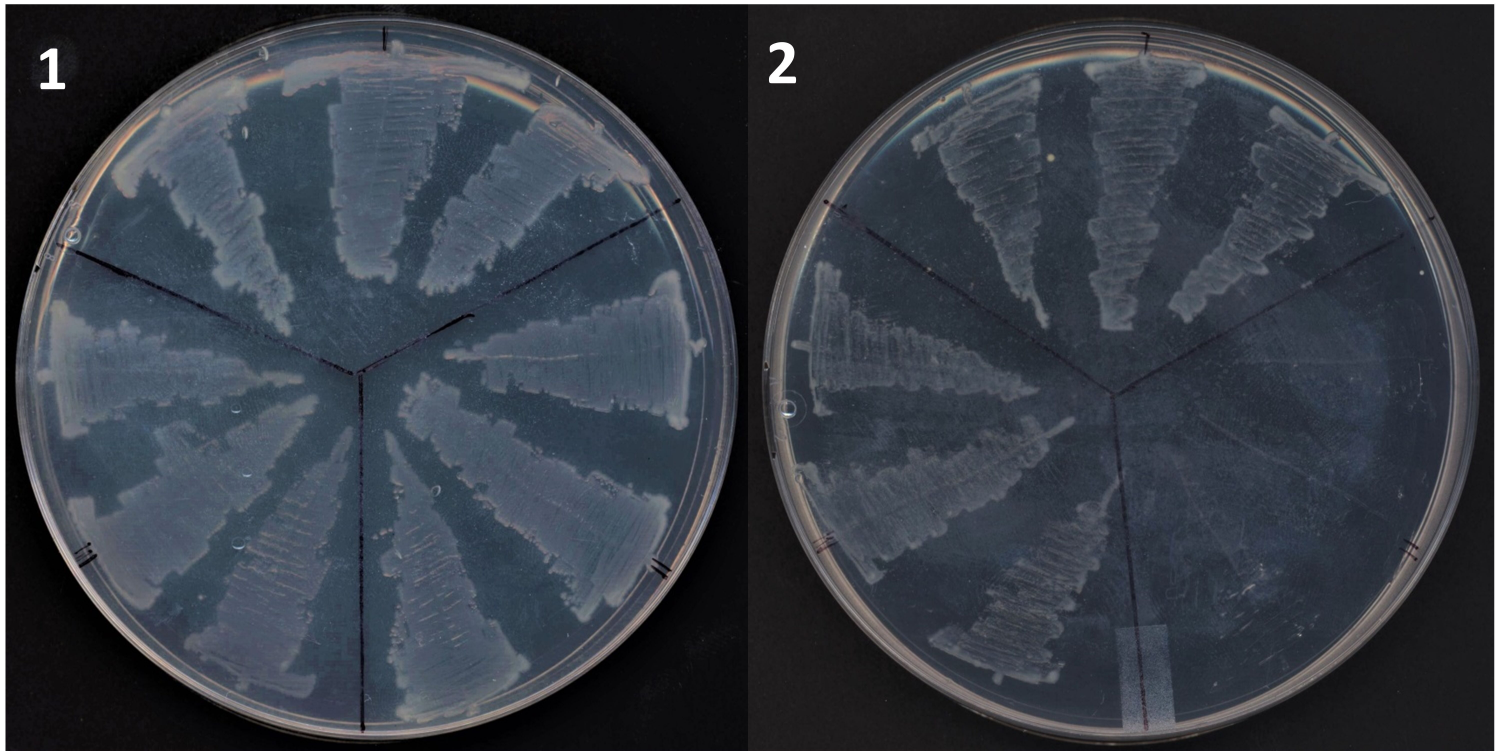


Figure 2

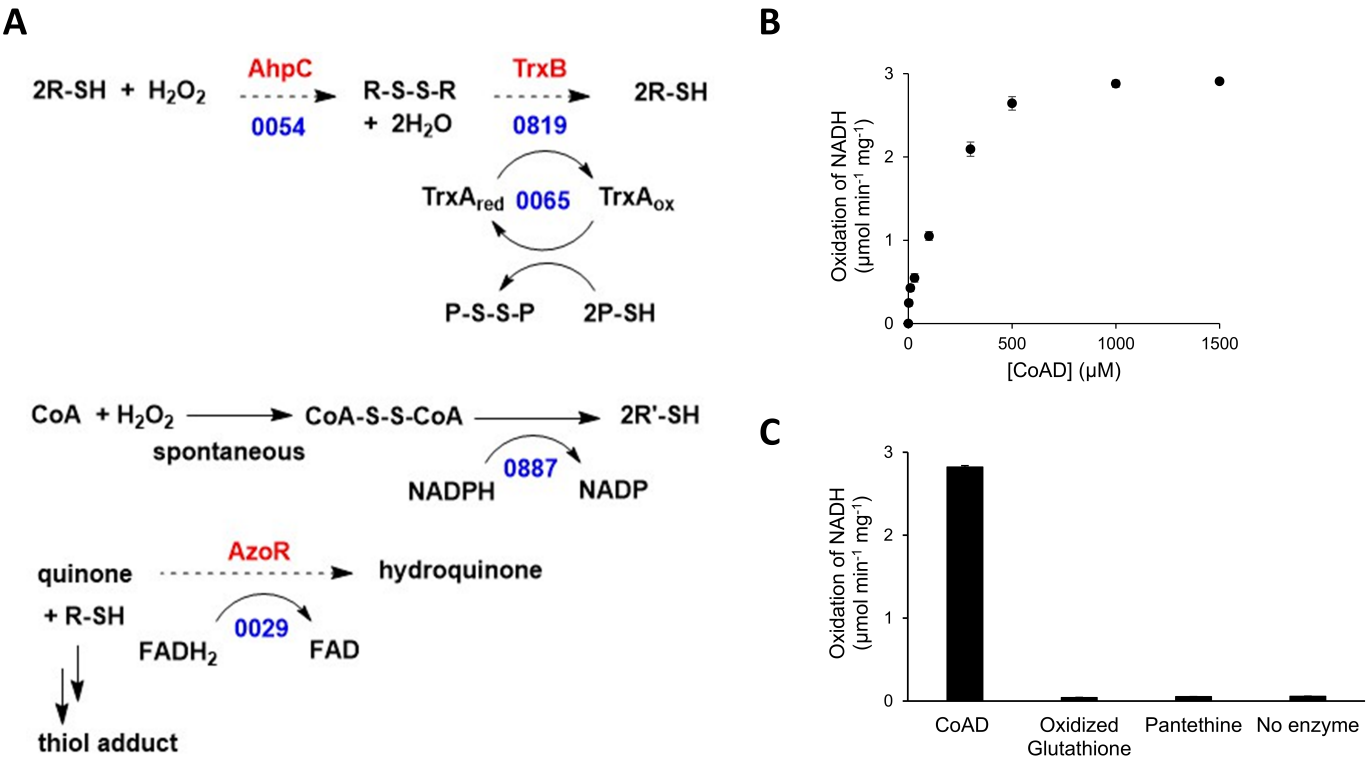
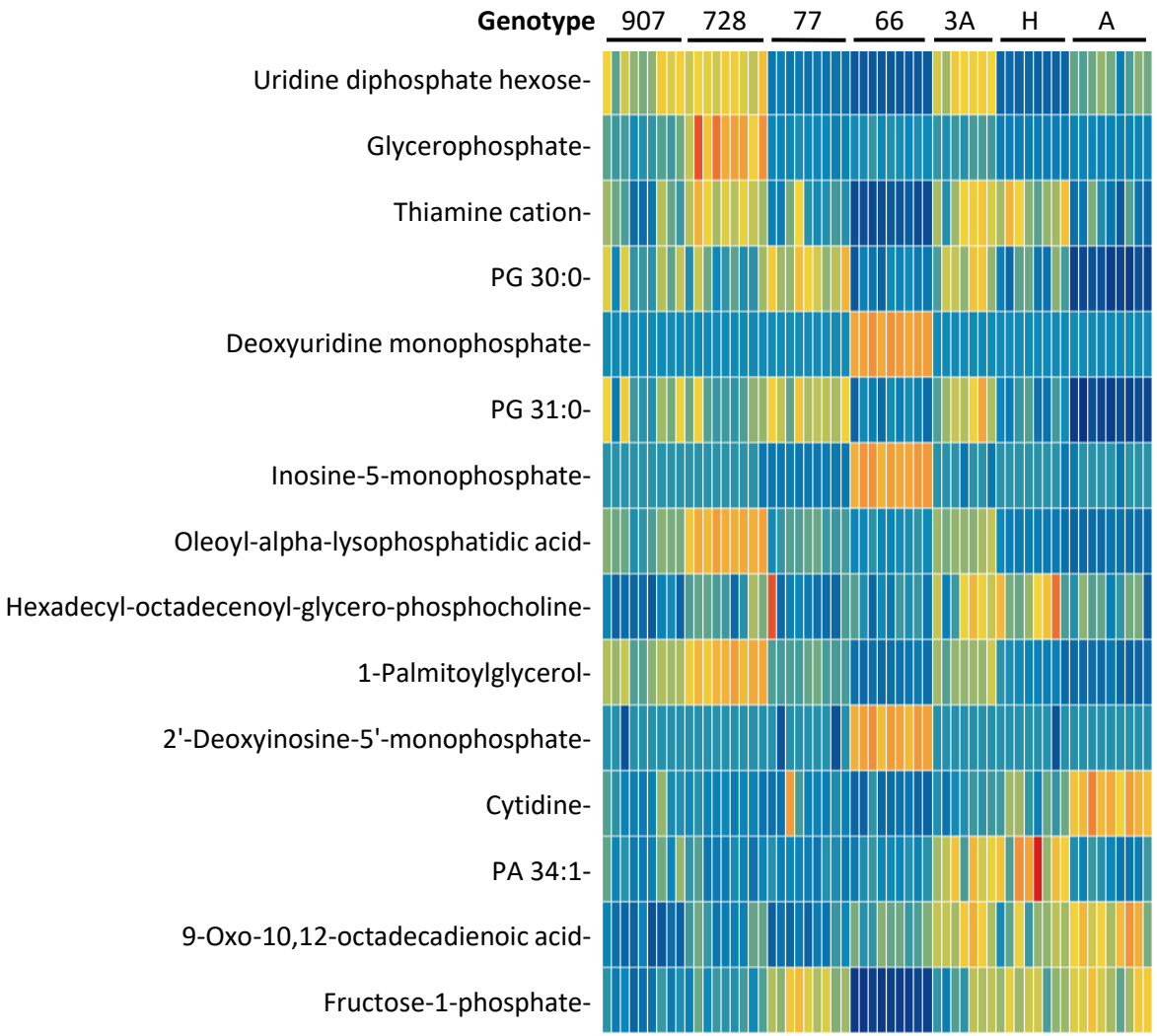

A
2R-SH + H2O2
AhpC
0054
R-S-S-R
+ 2H2O
TrxB
0819
2R-SH
TrxAred
0065
TrxAox
P-S-S-P
2P-SH
CoA + H2O2
spontaneous
CoA-S-S-CoA
2R'-SH
NADPH
0887
NADP
quinone
+ R-SH
AzoR
hydroquinone
FADH2
0029
FAD
thiol adduct
B
Oxidation of NADH
(µmol min-1 mg-1)
[CoAD] (µM)
0
500
1000
1500
1
2
3
C
Oxidation of NADH
(µmol min-1 mg-1)
CoAD
Oxidized
Glutathione
Pantethine
No enzyme

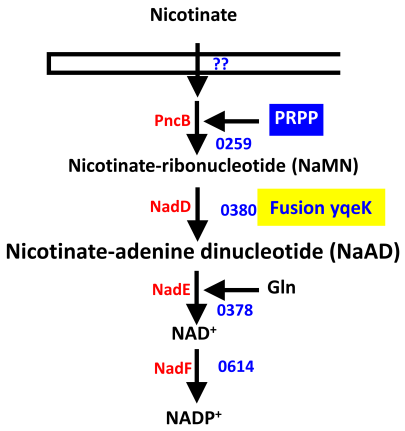
Figure 3



Label	genotype
_907	JCVI-syn3A_Δ907
_728	JCVI-syn3A_Δ728
_77	JCVI-syn3A_Δ077
_66	JCVI-syn3A_Δ066
_3A	JCVI-syn3A
_H	JCVI-syn3A_H230H
_A	JCVI-syn3A_H230A

Figure 4

A


Nicotinate
??
PncB
PRPP
0259
Nicotinate-ribonucleotide (NaMN)
NadD
0380
Fusion yqeK
Nicotinate-adenine dinucleotide (NaAD)
NadE
Gln
0378
NAD+
NadF
0614
NADP+

B

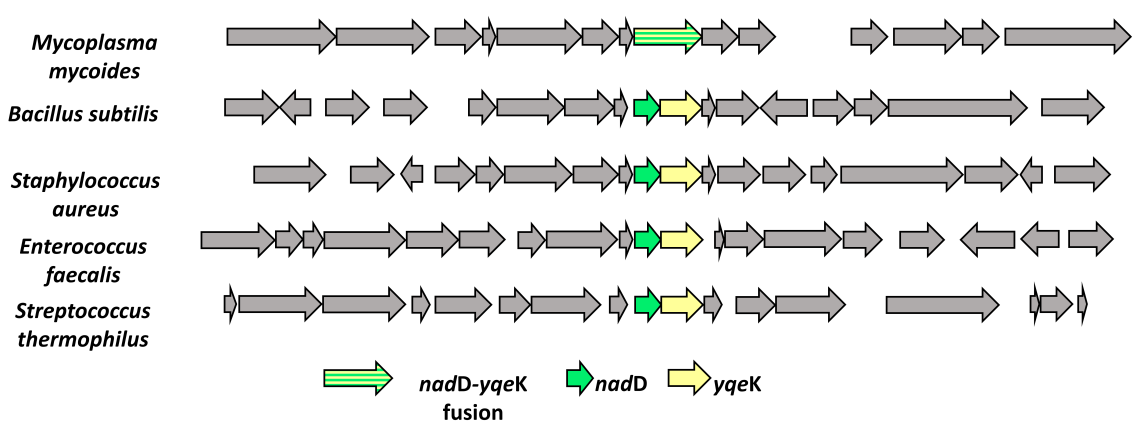

Mycoplasma mycoides
Bacillus subtilis
Staphylococcus aureus
Enterococcus faecalis
Streptococcus thermophilus
nadD-yqeK fusion
nadD
yqeK

Figure 5

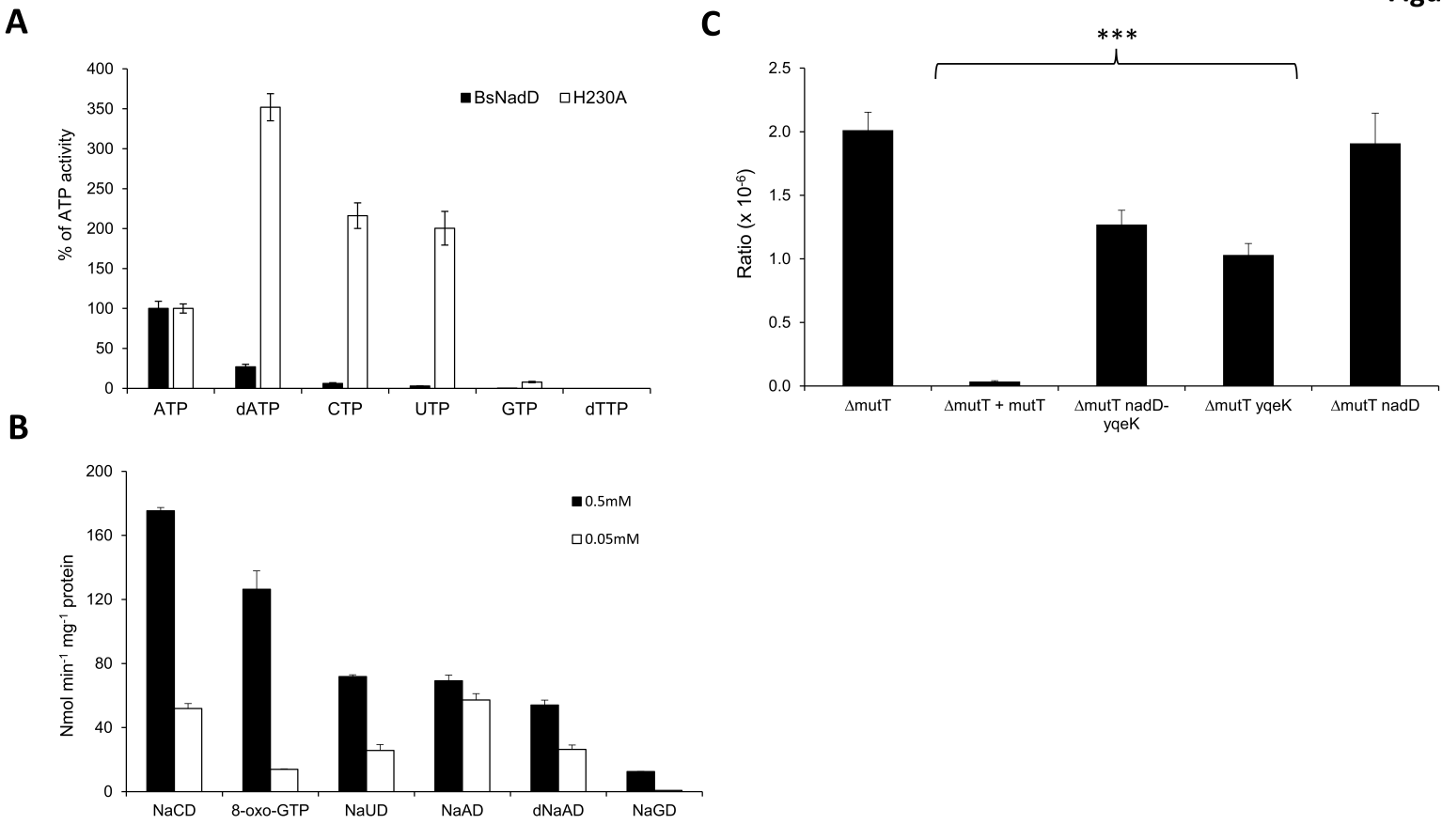

A
400
350
300
250
200
150
100
50
0
% of ATP activity
BsNadD
H230A
ATP
dATP
CTP
UTP
GTP
dTTP
B
200
160
120
80
40
0
Nmol min-1 mg-1 protein
0.5mM
0.05mM
NaCD
8-oxo-GTP
NaUD
NaAD
dNaAD
NaGD
C
***
2.5
2.0
1.5
1.0
0.5
0.0
Ratio (x 10-6)
ΔmutT
ΔmutT + mutT
ΔmutT nadD-yqeK
ΔmutT yqeK
ΔmutT nadD

Figure 6

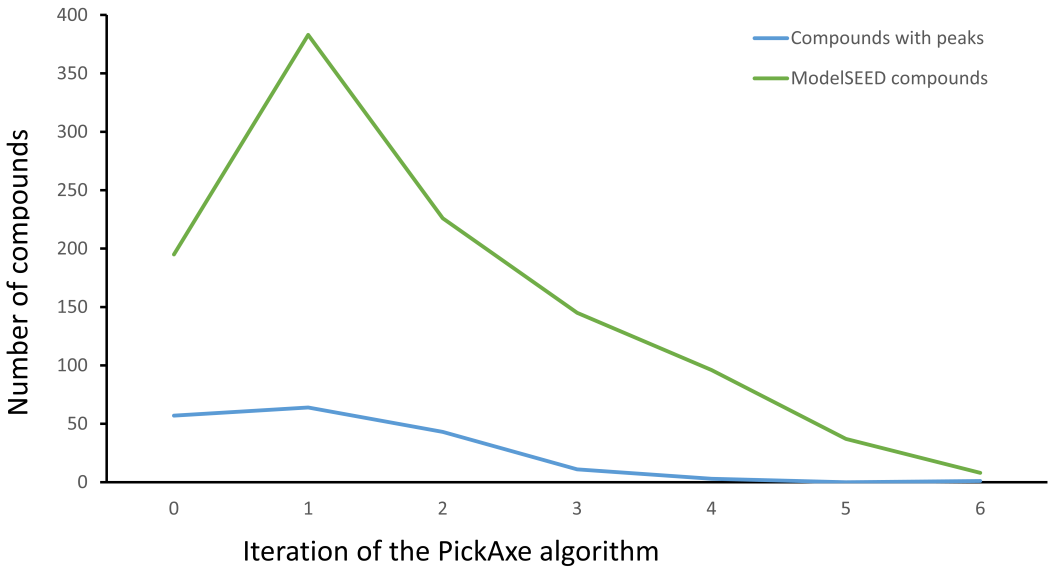

Compounds with peaks
ModelSEED compounds
Number of compounds
400
350
300
250
200
150
100
50
0
0
1
2
3
4
5
6
Iteration of the PickAxe algorithm

Figure 7

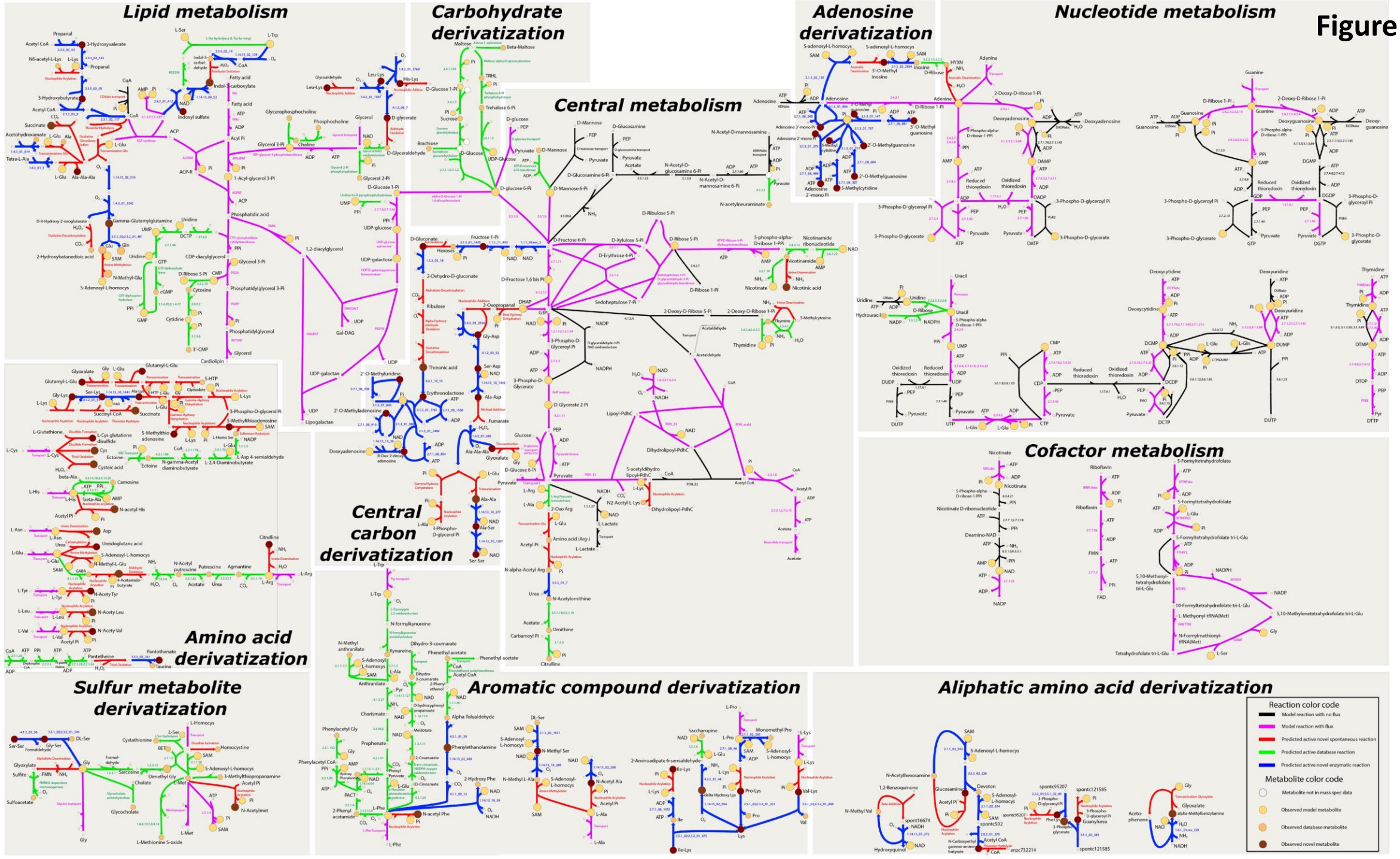
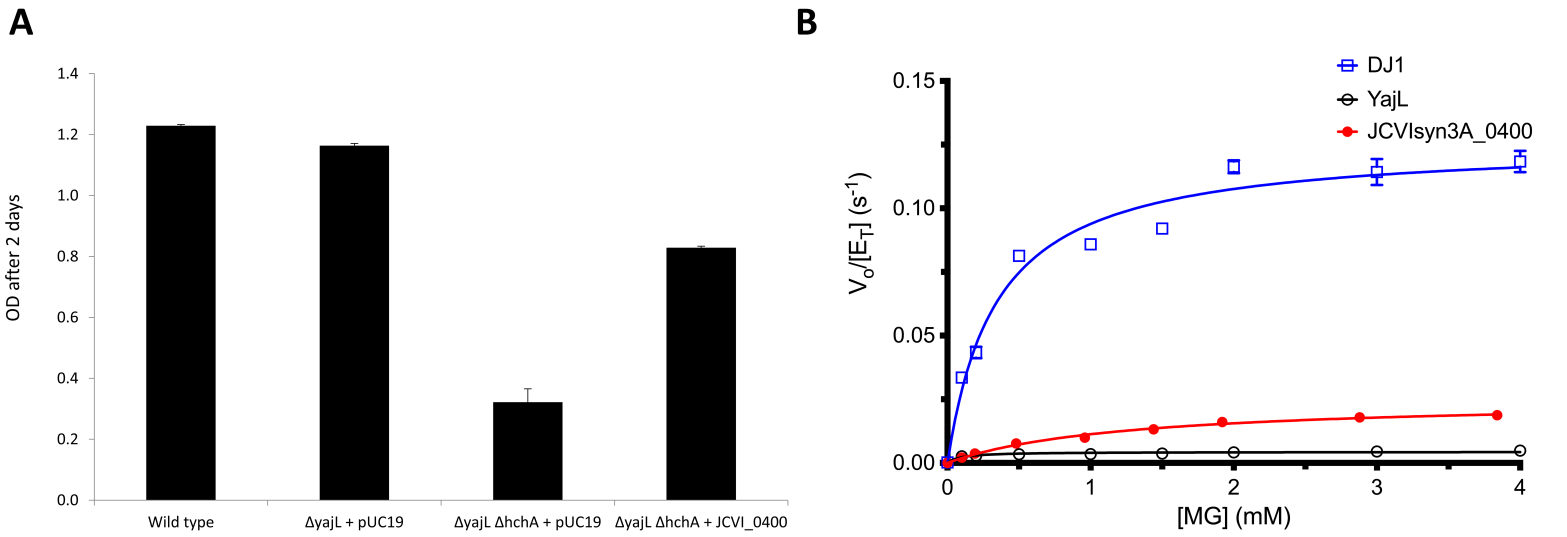

Lipid metabolism
Carbohydrate derivatization
Adenosine derivatization
Nucleotide metabolism
Central metabolism
Central carbon derivatization
Cofactor metabolism
Amino acid derivatization
Sulfur metabolite derivatization
Aromatic compound derivatization
Aliphatic amino acid derivatization
Reaction color code
Model reaction with no flux
Model reaction with flux
Predicted active novel spontaneous reaction
Predicted active database reaction
Predicted active novel enzymatic reaction
Metabolite color code
Metabolite not in mass spec data
Observed model metabolite
Observed database metabolite
Observed novel metabolite

Figure 8


A
OD after 2 days
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Wild type
ΔyajL + pUC19
ΔyajL ΔhchA + pUC19
ΔyajL ΔhchA + JCVI_0400
B
DJ1
YajL
JCVIsyn3A_0400
$V_o/[E_T]$ (s^{-1})
0.15
0.10
0.05
0.00
0
1
2
3
4
[MG] (mM)