Engineering and Evolution of *Saccharomyces cerevisiae* for Synthetic Formatotrophic Growth via the Reductive Glycine Pathway

Viswanada Reddy Bysani Kondagari

Max-Planck-Institut
für Molekulare Pflanzenphysiologie

Dissertation
zur Erlangung des akademischen Grades “doctor rerum naturalium” (Dr. rer. nat.)

in der Wissenschaftsdisziplin
“Systemische und Synthetische Biologie”

eingereicht an der Mathematisch-Naturwissenschaftlichen Fakultät der Universität Potsdam am 29th August 2022

Supervisors

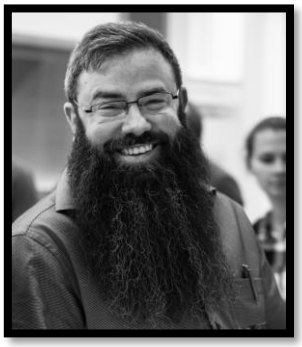
Prof. Dr. Dr. h.c. Mark Stitt em
Dr. habil. Arren Bar-Even (Late)
Dr. Fabian Machens

Thesis Reviewers

Prof. Dr. RA (Ruud) Weusthuis
Prof. Dr. Lars M. Blank
Prof. Dr. Dr. h.c. Mark Stitt em

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-58222
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-582222

Dedicated to



Dr. Arren Bar-Even
1980-2020

Declaration of Work

DECLARATION OF WORK

The work described in this dissertation, submitted to the Institute of Biochemistry and Biology at the University of Potsdam, has been carried out between July 2018 and June 2022 in the research group of Dr. Arren Bar-Even at the Max-Planck-Institute of Molecular Plant Physiology, Germany. Apart from my supervisor's guidance, or as clearly stated, the content of this thesis is the result of work completed by myself. I declare that this thesis has not been submitted to another examining body, either partially or wholly, as part of a doctoral degree. The thesis as a whole has not been published or submitted for publication.

EIDESSTATTLICHE ERKLÄRUNG

Die vorliegende Dissertationsschrift ist das Ergebnis meiner eigenen praktischen Arbeit. Sie wurde von Juli 2018 bis Juni 2022, am Max-Planck-Institut für molekulare Pflanzenphysiologie in Potsdam, in der Arbeitsgruppe von Dr. Arren Bar-Even durchgeführt. Ich erkläre, dass ich die vorliegende Arbeit selbstständig und ohne unerlaubte Hilfe angefertigt habe. Es wurden keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Die verwendeten Quellen sind als solche im Text kenntlich gemacht. Die Arbeit wurde zuvor noch an keiner anderen Stelle eingereicht.

Potsdam, 29-08-2022

Viswanada Reddy Bysani Kondagari

Table of content

List of tables

List of figures

List of supplementary information

List of plasmid maps S1-S8

List of figures S1-S12

List of tables S1-S6

Abbreviations

GHG : Green House Gases

UNFCCC : United Nations Framework Convention on Climate Change

IPCC : Intergovernmental Panel on Climate Change

RGP : Reductive Glycine Pathway

rAcCoAP : Reductive Acetyl-CoA pathway

GCV : Glycine Cleavage System

GDC : Glycine Decarboxylase Complex (=GCV)

GSC : Glycine Synthase Complex (Reversal of GCV)

C1 : One-carbon

CHO-THF : 10-Formyltetrahydrofolate

CH-THF : 5,10-Methenyltetrahydrofolate

CH_2-THF : 5,10-Methylenetetrahydrofolate

CH_3-THF : 5-Methyltetrahydrofolate

C1 unites : CHO-THF, CH-THF, CH_2-THF, and CH_3-THF

3-PGA : 3-phosphoglyceric acid

SHMT : Serine hydroxymethyltransferase

ΔS : Serine biosensor

Ev : Evolution round

ALE : Adaptive Laboratory Evolution

NGS : Next Generation Sequencing

CDS : Coding sequence

ARS : Autonomously replicating sequence

1. Summary

Increasing demand for food, healthcare, and transportation arising from the growing world population is accompanied by and driving global warming challenges due to the rise of the atmospheric CO_2 concentration. Industrialization for human needs has been increasingly releasing CO_2 into the atmosphere for the last century or more. In recent years, the possibility of recycling CO_2 to stabilize the atmospheric CO_2 concentration and combat rising temperatures has gained attention. Thus, using CO_2 as the feedstock to address future world demands is the ultimate solution while controlling the rapid climate change. Valorizing CO_2 to produce activated and stable one-carbon feedstocks like formate and methanol and further upgrading them to industrial microbial processes to replace unsustainable feedstocks would be crucial for a future biobased circular economy. However, not all microbes can grow on formate as a feedstock, and those microbes that can grow are not well established for industrial processes.

S. cerevisiae is one of the industrially well-established microbes, and it is a significant contributor to bioprocess industries. However, it cannot grow on formate as a sole carbon and energy source. Thus, engineering *S. cerevisiae* to grow on formate could potentially pave the way to sustainable biomass and value-added chemicals production.

The Reductive Glycine Pathway (RGP), designed as the aerobic twin of the anaerobic Reductive Acetyl-CoA pathway, is an efficient formate and CO_2 assimilation pathway. The RGP comprises of the glycine synthesis module (Mis1p, Gcv1p, Gcv2p, Gcv3p, and Lpd1p), the glycine to serine conversion module (Shmtp), the pyruvate synthesis module (Cha1p), and the energy supply module (Fdh1p). The RGP requires formate and elevated CO_2 levels to operate the glycine synthesis module. In this study, I established the RGP in the yeast system using growth-coupled selection strategies to achieve formate and CO_2-dependent biomass formation in aerobic conditions.

Firstly, I constructed serine biosensor strains by disrupting the native serine and glycine biosynthesis routes in the prototrophic S288c and FL100 yeast strains and insulated serine, glycine, and one-carbon metabolism from the central metabolic network. These strains cannot grow on glucose as the sole carbon source but require the supply of serine or glycine to complement the engineered auxotrophies. Using growth as a readout, I employed these strains as selection hosts to establish the RGP. Initially, to achieve this, I engineered different serine-hydroxymethyltransferases in the genome of serine biosensor strains for efficient glycine to serine conversion. Then, I implemented the glycine synthesis module of the RGP in these strains for the glycine and serine synthesis from formate and CO_2. I successfully conducted Adaptive Laboratory Evolution (ALE) using these strains, which yielded a strain

capable of glycine and serine biosynthesis from formate and CO_2. Significant growth improvements from 0.0041 h^{-1} to 0.03695 h^{-1} were observed during ALE. To validate glycine and serine synthesis, I conducted carbon tracing experiments with ^{13}C formate and $^{13}CO_2$, confirming that more than 90% of glycine and serine biosynthesis in the evolved strains occurs via the RGP. Interestingly, labeling data also revealed that 10-15% of alanine was labelled, indicating pyruvate synthesis from the formate-derived serine using native serine deaminase (Cha1p) activity. Thus, RGP contributes to a small pyruvate pool which is converted to alanine without any selection pressure for pyruvate synthesis from formate. Hence, this data confirms the activity of all three modules of RGP even in the presence of glucose. Further, ALE in glucose limiting conditions did not improve pyruvate flux via the RGP.

Growth characterization of these strains showed that the best growth rates were achieved in formate concentrations between 25 mM to 300 mM. Optimum growth required 5% CO_2, and dropped when the CO_2 concentration was reduced from 5% to 2.5%.

Whole-genome sequencing of these evolved strains revealed mutations in genes that encode *Gdh1p*, *Pet9p*, and *Idh1p*. These enzymes might influence intracellular NADPH, ATP, and NADH levels, indicating adjustment to meet the energy demand of the RGP. I reverse-engineered the *GDH1* truncation mutation on unevolved serine biosensor strains and reproduced formate dependent growth. To elucidate the effect of the *GDH1* mutation on formate assimilation, I reintroduced this mutation in the S288c strain and conducted carbon-tracing experiments to compared formate assimilation between WT and Δ*gdh1* mutant strains. Comparatively, enhanced formate assimilation was recorded in the Δ*gdh1* mutant strain.

Although the ^{13}C carbon tracing experiments confirmed the activity of all three modules of the RGP, the overall pyruvate flux via the RGP might be limited by the supply of reducing power. Hence, in a different approach, I overexpressed the formate dehydrogenase (Fdh1p) for energy supply and serine deaminase (Cha1p) for active pyruvate synthesis in the S288c parental strain and established growth on formate and serine without glucose in the medium. Further reengineering and evolution of this strain with a consistent energy, and formate-derived serine supply for pyruvate synthesis, is essential to achieve complete formatotrophic growth in the yeast system.

ZUSAMMENFASSUNG

Die steigende Nachfrage nach Lebensmitteln, Gesundheitsfürsorge und Transportmitteln durch die stetig wachsende Weltbevölkerung, sorgen für eine dramatisch fortschreitende globale Erwärmung; verursacht durch den massiven weltweiten CO_2 Ausstoß. Durch die Industrialisierung wurde im vergangenen Jahrhundert immer mehr CO_2 in die Atmosphäre freigesetzt. Recycling von CO_2 hat in den letzten Jahren an Aufmerksamkeit gewonnen, um die steigenden Temperaturen zu stabilisieren. CO_2 ist somit der ultimative Rohstoff, um den künftigen weltweiten Bedarf zu decken und gleichzeitig den raschen Klimawandel zu kontrollieren. Die Verwendung von CO_2 zur Herstellung von aktivierten Ein-Kohlenstoff-Verbindungen wie Formiat und Methanol und die weitere Aufwertung durch mikrobielle Prozesse, wäre für die biobasierte Kreislaufwirtschaft von entscheidender Bedeutung. Allerdings können die allermeisten Mikroorganismen auf den genannten Ein-Kohlenstoffverbindungen- als Ausgangsstoff nicht wachsen und diejenigen, dies es können sind für industrielle Prozesse nicht geeignet.

S. cerevisiae gehört zu den industriell etablierten Mikroorganismen und leistet einen wichtigen Beitrag zur Bioprozessindustrie. Sie kann jedoch nicht auf Formiat als einziger Kohlenstoff- und Energiequelle wachsen. Die biotechnologische Anpassung von *S. cerevisiae* für das Wachstum auf Formiat könnte daher nachhaltige Wege für die Produktion von Biomasse und wertschöpfenden Chemikalien ebnen.

Der Reduktive Glycinweg (RGP), der als aerober Zwilling des anaeroben Reduktiven Acetyl-CoA-Wegs konzipiert wurde, ist ein effizienter Formiat- und CO_2 Assimilationsweg. Der RGP besteht aus dem Glycin-Synthesemodul (Mis1p, Gcv1p, Gcv2p, Gcv3p und Lpd1p), dem Modul zur Umwandlung von Glycin in Serin (Shmtp), dem Pyruvat-Synthesemodul (Cha1p) und dem Energieversorgungsmodul (Fdh1p). Der RGP benötigt Formiat und erhöhte CO_2 Verfügbarkeit, um das Glycin-Synthesemodul zu betreiben. In dieser Studie habe ich das RGP im Hefesystem mit wachstumsgekoppelten Selektionsstrategien etabliert, um ein von Formiat und CO_2 abhängiges zelluläres Wachstum unter aeroben Bedingungen zu erreichen.

Zunächst habe ich Serin-Biosensor-Stämme konstruiert, indem ich die nativen Serin- und Glycin-Biosynthesewege in den prototrophen Stämmen S288c und FL100 unterbrochen und den Serin-, Glycin- und Ein-Kohlenstoff-Stoffwechsel vom zentralen Stoffwechselnetz isoliert habe. Diese Stämme können nicht mit Glukose als alleinige Kohlenstoffquelle wachsen, sondern benötigen die Zufuhr von Serin oder Glycin, um die eingeführten Auxotrophien zu ergänzen. Unter Verwendung von zellulärem Wachstum als Indikator, habe ich diese Stämme als Selektionswirte verwendet, um den RGP zu etablieren. Zu diesem Zweck habe ich durch genomische Integration von Serin-Hydroxymethyltransferasen (SHMTs) in diese Biosensorstämme effiziente Module zur Umwandlung von Glycin in Serin geschaffen. Dann implementierte ich das Glycin-Modul des RGP in die Stämme für die Glycin- und Serinsynthese aus Formiat und CO_2. Mit diesen Stämmen führte ich erfolgreich eine adaptive Laborevolution (ALE) durch, die einen Stamm hervorbrachte, der zur Glycin- und Serinbiosynthese aus Ameisensäure und CO_2 fähig ist. Während der ALE wurden signifikante Wachstumsverbesserungen von 0,0041 h^{-1} auf 0,03695 h^{-1} beobachtet. Um die Glycin- und

Serinsynthese zu bestätigen, führte ich Experimente zur Kohlenstoffverfolgung mit ^{13}C-Formiat und $^{13}CO_2$ durch, die bestätigten, dass mehr als 90% der Glycin- und Serinbiosynthese über den RGP erfolgt. Interessanterweise ergaben die Markierungsdaten auch 10-15% markiertes Alanin, was auf eine Pyruvatsynthese aus dem von Formiat abgeleiteten Serin durch die native Serindeaminase (Cha1p) hinweist. Somit trägt der RGP zu einem kleinen Pyruvat-Pool bei, ohne dass ein Selektionsdruck für die Pryuvat Synthese aus Formiat besteht. Diese Daten bestätigen somit die Aktivität aller drei Module von RGP auch in Gegenwart von Glukose. Weitere ALE unter glukoselimitierenden Bedingungen verbesserte den Pyruvatfluss aus dem RGP allerdings nicht.

Die Wachstumscharakterisierung der beschriebnen Stämme zeigte, dass die besten Wachstumsraten bei Formiatkonzentrationen zwischen 25 mM und 300 mM erzielt wurden. Für optimales Wachstum wurde 5% CO_2 benötigt und die Wachstumsrate verschlechterte sich bei einer CO_2 Konzentration von 2.5 %.

Die Sequenzierung des gesamten Genoms dieser weiterentwickelten Stämme ergab Mutationen in Genen, die für *Gdh1p*, *Pet9p* und *Idh1p* kodieren. Diese Enzyme beeinflussen den intrazellulären NADPH-, ATP- und NADH-Spiegel, was auf den Energiebedarf für die Aktivität des RGP hinweist. Ich habe die GDH1-Mutation in nicht evolvierte Serin-Biosensor-Stämme eingebracht und damit das von Formiat und CO_2 abhängige Wachstum reproduziert. Um die Auswirkung der GDH1-Mutation auf die Formationsassimilation zu klären, habe ich diese Mutation in den WT-Stamm eingeführt und ein Experiment zur Kohlenstoffverfolgung mit ^{13}C-Formiat und Glukose durchgeführt. Es wurde nachgewiesen, dass die *gdh1*-Mutante im Vergleich zum WT-Stamm eine verbesserte Assimilation aufweist.

Obwohl die ^{13}C-Isotopenmarkierung die Aktivität aller drei Module der RGP bestätigte, könnte der Gesamtpyruvatfluss über den RGP durch die Versorgung mit Redox-Äuquivalenten begrenzt sein. In einem anderen Ansatz wurden daher die Formiatdehydrogenase (Fdh1p) für die Energieversorgung und die Serindesaminase (Cha1p) für die aktive Pyruvatsynthese überexprimiert und ein Wachstum mit Formiat und Serin ohne Glukose in der WT-Hefe festgestellt. Die weitere Entwicklung des gesamten Stoffwechsels in Kombination mit Evolutionsstrategien,und einer aus Formiat gewonnenen Serin- und Energiezufuhr für die Pyruvatsynthese ist unerlässlich, um ein vollständiges formatotrophes Wachstum im Hefesystem zu erreichen.

2. Introduction

Current society mainly depends on fossil fuels and coal to produce commodity products and energy to meet human needs. The growing world population is creating an increased requirement for fuel for heating, cooling, transportation, food, animal feed, and medicines, which places an increased burden on fossil fuel-based industrial processes. Continuous exploitation of fossil fuels is not sustainable; the reserves are anyway finite and it results in the constant release of Green House Gasses (GHG) into the atmosphere. These issues drive industries to implement alternative sustainable biotechnology-based refineries to produce food, valuable chemicals, and fuel using sustainable feedstocks. Most biotechnology-based production processes currently use sugars, starch, molasses, and lignocellulosic biomass[1,2]. Some of these feedstocks undermine food security, e.g., sugars and starch. Other feedstocks are not genuinely sustainable[2], as processing and use of them lead to GHG generation contributing to climate change. Hence, alternative sustainable feedstock supplies are essential for industrial production pipelines to address future world requirements while tackling climate change essentially via the circularization of industrial carbon emissions.

2.1. Circular carbon bio-economy – the reuse of the carbon emissions

Increased industrial processes led to the release of CO_2 since the industrial revolution, leading to global warming[3]. The rise of the earth's surface temperature is causing calamities worldwide (IPCC 2014). Scientific evidence shows that global warming and associated climate change consequences are due to the continuous release of GHG i.e. CO_2, CH_4, and N_2O (nitrous oxide) by human activities[4,5]. Accumulation of these gasses in the lower layer of the atmosphere absorbs the radiation reflected from the earth's surface, which in turn increases atmospheric temperature causing the so-called greenhouse effect[6]. These climate change consequences have become visible and this is an urgent issue on the international stage. As a response, Intergovernmental Panel on Climate Change (IPCC) was formed by the United Nations (UN), which summarizes scientific information to the UN body on Climate change[7]. The IPCC reports are critical for international agreements like the Kyoto Protocol or Paris Agreements, or COP26 (UNFCCC) that aim to reduce global GHG emissions. GHG emissions need to be reduced dramatically to prevent the rise of the global average temperature from rising more than 2 ºC above pre-industrial levels. CO_2 is the primary GHG contributor (IPCC 2013), and reduction of its emissions could mitigate global warming. According to the IPCC estimation, to stabilize global warming at 1.5 ºC above pre-industrial levels, anthropogenic CO_2 release must be reduced by 45% from 2010 to 2030 and reach a net-zero around 2050. To achieve this, solutions with multiple levels of interconnectivity across the global

communities are required (IPCC 2018). Different ways to reduce the concentration of CO_2 in the atmosphere include more efficient energy usage and the use of more sustainable alternative energy sources, as well as CO_2 capture and sequestration (CCS)[8]. CCS involves CO_2 capture, concentration, transportation, and storage. Among the several technological options, it is expected to play a significant role in recycling CO_2 emissions[8,9]. CCS's current concept for CO_2 capture is at a point source (e.g., a cement factory or power plant) and injection of liquefied CO_2 into suitable geological formations for long-term storage[10].

Several concepts to reuse the captured CO_2 have been in discussion and execution recently, including pumping CO_2 back in underground and locking up CO_2 for the shorter or longer term. Storing CO_2 into concrete and using it for the global demand of building material supply could be scalable and has the potential to dominate the CO_2 conversion market as one of the long-term storage strategies[11–13]. Thus, CO_2 directly or indirectly can be used as a feedstock in a circular carbon economy. Model studies were conducted that can be used to select appropriate sources, capture technologies, transportation networks, and CO_2 storage sites for the nationwide reduction of CO_2 in the Netherlands. The cost is estimated as a minimum of € 47.8 billion for 25 years of operation and capture of 54 Mtpa (million tons per annum) CO_2[14].

Several chemical processes use CO_2 to produce urea, polycarbonates, and hydrocarbons. While being already commercialized, most chemical solutions suffer from low product selectivity and a limited product spectrum[15,16]. Roughly, 200 million tons of CO_2 are being used for urea production; however, urea production and urea utilization release CO_2 back to nature quickly; hence, CO_2 is not stored for the long-term in this case. Studies have been made on the electrochemical reduction of CO_2 to methanol and other fuel chemicals[17–20]. Several strategies are presented for the bio-electrochemical reduction of CO_2[21]. Many scientific studies have been proposed and conducted on CO_2 reduction, e.g., syngas ($CO+H_2$) production from CO_2 is already being tested at the pilot plant level[22]. Scientific studies were conducted on using captured CO_2 as energy storage material, e.g., upcycling CO_2 and renewable hydrogen and storing them as methanol allow the cyclic use of carbon as storage material[23]. Methanol production from captured CO_2 is closer to commercialization, as industries are being formed for the production of sustainable methanol via CO_2 reduction[13]. Methane, another abundant GHG gas, can be used for methanol production, which can be used as a feedstock for methylotrophic yeast and bacteria[24–26]. Formate is an attractive alternative molecule that is used as a precursor for chemical production, and can be produced via electrochemical reduction of CO_2[27]. Formate that is produced via the direct reduction of CO_2 from solar energy is also been investigated to use as a fuel cell for power supply[28]. Recent reports say formate production from CO_2 reduction is expected to be tested next year on a pilot scale[22].

2.2. One-carbon feedstock for biotechnology processes

Autotrophic organisms fix inorganic carbon into organic compounds and are the main gateway between the inorganic and living worlds. Autotrophs dominate biomass production on Earth[29,30], supplying all of our food and - partly directly but mainly indirectly - most of our fuel. Recent studies successfully converted heterotrophic *E. coli* and a methylotrophic yeast to produce their complete biomass from CO_2[31,32]. Thus, the captured CO_2 can be used as a feedstock in biotechnological applications where autotrophic, acetogenic, formatotrophic or methylotrophic microorganisms could feed on it directly or indirectly[33–35]. Biological systems operate under mild conditions, excel at product specificity, and achieve economic viability in decentralized production facilities at a smaller scale[24,36]. However, gaseous C1 feedstocks, E.g., CO_2, CO, CH_4, and CO_2/H_2, are limited by their solubility in water, leading to mass transfer constraints and resulting in less microbial growth. In contrast, non-gaseous C1 compounds like formate and methanol bypass the mass transfer barrier and can potentially support higher microbial productivities[33,35]. Another significant advantage of these two compounds is that, unlike the C1 gases, they can be easily stored and transported. This enables a spatial and temporal decoupling of the abiotic feedstock production from the biotic feedstock consumption, an important feature that serves to insulate microbial growth and bioproduction from the production of C1 feedstocks[35]. Many scientific and commercial studies were conducted on the microbial cultivation on methanol[37], and industrial production of microbial proteins from methanol was pursued to provide human and animal nutrition[1]. However, until recently, the methanol was always produced from fossil fuels rather than CO_2. In contrast to methanol, formate has mainly been neglected as a potential industrial feedstock although it was used as an auxiliary carbon source to supplement the cell with reducing power and thus boost bioproduction[38].

Several ways were reported for formate production industrially; by hydrogenation of CO_2, photoreduction of CO_2 using solar energy[28,39–41], oxidation of natural gas and biomass (waste plant material, woody tissue, and even microorganisms), and hydration (water addition) of syngas[42,41]. Among the possible mediator compounds, hydrogen, carbon monoxide, and formate can be produced at high efficiency and rate[43]. Hence, formate can also be used as hydrogen storage, thus it is also considered as an alternative for a hydrogen-based economy[23,44]. Formate can be produced by simple electrochemical reduction of CO_2 in the presence of water[45]. Formate synthesis through the electrochemical reduction of CO_2 boasts Faradaic efficiency of >90%[46]. With further developments in electrochemical, photochemical, and catalytic methods of generating formate and potential scale-up possibilities, interest in its use as a microbial feedstock is rising[17,22,25,27,45].

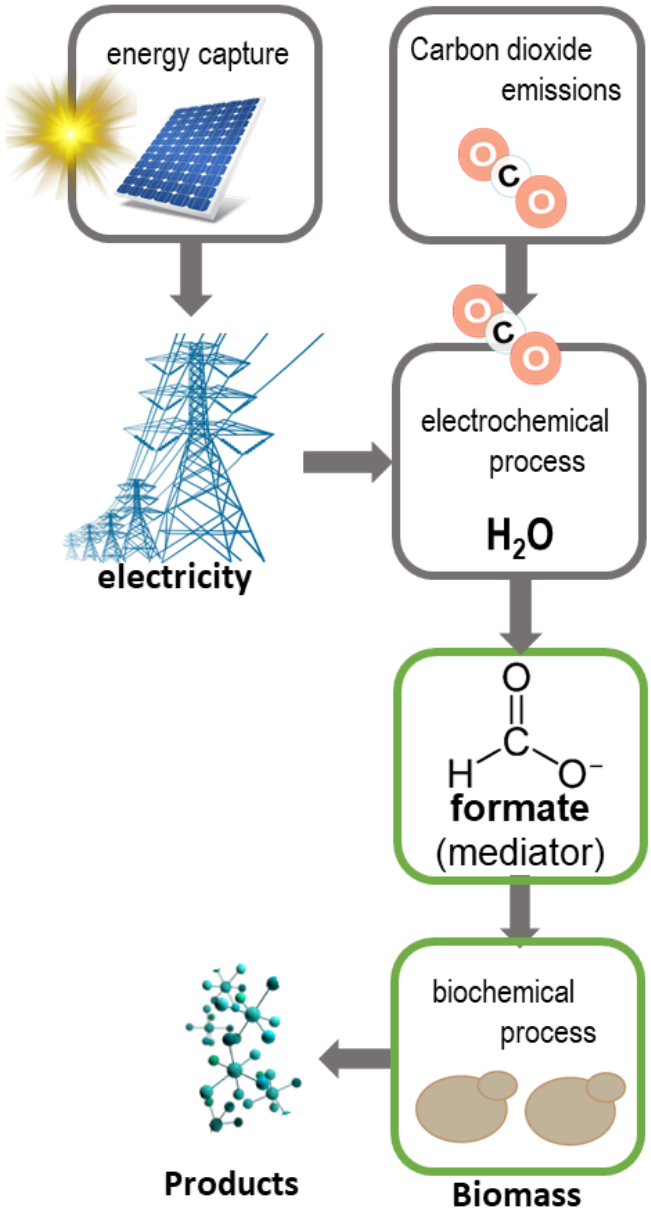


Figure 1: **Simplified concept of formate mediated circular bio-economy**
The electrochemical reduction process of captured CO_2 using sustainable electricity for the production of the mediator molecule formate. Using formate as the feedstock in the biochemical processes with engineered *S. cerevisiae* to produce biomass and other valuable compounds.

Currently, glucose is widely used as feedstock in biorefineries; glucose production capacity is however limited and mainly through agriculture, while the main components needed for formate synthesis are electricity and CO_2, therefore, formate is a more promising candidate than glucose for displacing fossil carbon fuels[33]. Hence, formate is a potential feedstock that can replace sugar and starch like agriculture-based feedstocks for bioprocess industries for the production of biomass and valuable products. Thus, the formate mediated bio-economy is put forward as a potential efficient process at the heart of the circular carbon economy (**Figure 1**).

2.3. Formate as a feedstock for microbial cultivation

There are several microbes, primarily acetogen, methanotrophs, and methylotrophs, that can grow on one-carbon feedstocks like formate, methane, and methanol as the sole carbon and energy sources[35]. *Cupriavidus necator (formerly known as Ralstonia eutropha)* was explored for the production of higher alcohols using formate as the feedstock that is produced from CO_2 reduction using electricity[47]. The Calvin Benson Cycle is known to support growth on methanol and formate in multiple microbial lineages. In fact, aerobic growth on formate for biotechnological purposes was mostly explored with microorganisms that employ the Calvin Benson Cycle, mainly *C. necator* [48,49]. However, the use of the Calvin Benson Cycle for growth on either methanol or formate is characterized by a low energetic efficiency[35]. Many other acetogenic bacteria grow on formate, but the metabolism of these microbes is yet to be understood completely; however, these microbes are not well-established as industrial hosts due to difficulties in cultivation and engineering[50,51].

Unfortunately, most biotechnologically relevant microbes cannot utilize formate as a carbon source, although some can utilize formate as an energy source. Hence, engineering a synthetic metabolic pathway for formate assimilation in the model microbes to allow them to utilize formate as both carbon and energy source, has the potential for biotechnology-based production[52,53]. Recently, *E. coli* has been engineered and further evolved to use formate as a carbon and energy source[54]. This was achieved via the Reductive Glycine Pathway (RGP). In the same study, synthetic methylotrophy, via formate assimilation, was established by coupling the RGP with the Mdhp (methanol dehydrogenase) module[54]. In another study, the RGP was established in *C. necator*, and successfully established formatotrophic growth through RGP by replacing formatotrophic growth via the Calvin Benson cycle[55]. In a recent study, *Desulfovibrio desulfuricans*, a sulfate-reducing bacteria, is reported to grow using formate and CO_2 through its native RGP[56]. A metagenomics study revealed that the *Candidatus Phosphitivorax (Ca. P.) anaerolimi*, a phosphite (HPO_3^{2-})-oxidizing bacterium, uses CO_2 as an electron acceptor and transfers an electron from HPO_3^{2-} to CO_2 to produce formate which is further assimilated for growth/biomass production[57].

2.4. Architecture of the RGP for formate assimilation

There are >70 known metabolic reactions in which formate participates as a reactant. Based on this cumulative information, several synthetic and natural formate assimilation pathways were analyzed for their energy efficiencies[53,58]. Among several natural formate assimilation pathways, the Reductive Acetyl-CoA pathway (rAcCoAP) is the most ATP efficient (**Figure 2**),

but this pathway is restricted to anaerobic conditions due to the oxygen sensitivity of the enzymes involved.

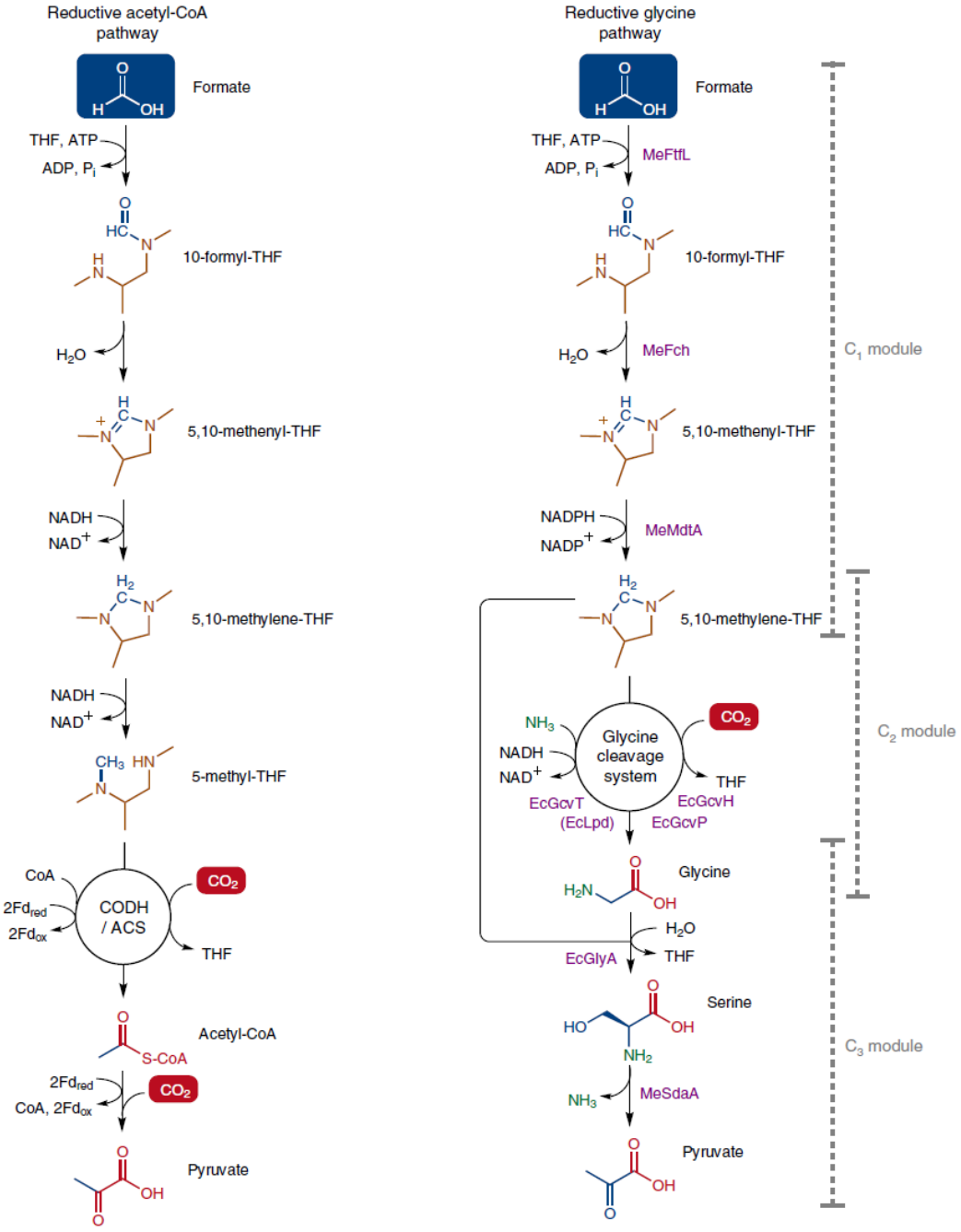


Figure 2: Basic structure of the RGP
The synthetic RGP is similar in structure to the rAcCoAP. Yet, while the latter pathway is restricted to anaerobic conditions, the former can operate under aerobic conditions. Both pathways are highly ATP-efficient, as only 1–2 ATP molecules are consumed in the conversion of formate to pyruvate (for example, instead of seven by the Calvin cycle). Me, *M. extorquens*; Ec, *E. coli*. Division of the pathway into modules is shown in gray to the right of the figure. This figure is adapted from Kim et al. (2020).

The Reductive Glycine Pathway (RGP) has been designed to be an aerobic twin pathway to the rAcCoAP with similar energy efficiency(Figure 2)[53,54]. Both pathways have limited overlap

with the central metabolic network. Both are linear pathways that start with the ligation of formate and tetrahydrofolate (THF) and proceed via reduction into a C1-THF intermediate, which is then condensed within an enzyme complex, with CO_2 to generate a two-carbon (C2) compound (acetyl-CoA or glycine). The C2 compound is finally condensed with another C1 moiety to generate pyruvate as the biomass precursor. In engineering it into the prokaryotic system, e.g., *E. coli*[54], the RGP was divided into four modules (1) a C1 module (C1M), consisting of formate-THF ligase, methenyl-THF cyclohydrolase, and methylene-THF dehydrogenase. All the genes were from *Methylobacterium extorquens*[59]; (2) a C2 module (C2M), consisting of the endogenous enzymes of the GSC (GcvT, GcvH, and GcvP); (3) a three-carbon (C3) module (C3M), consisting of serine hydroxymethyltransferase (SHMT) and serine deaminase; and (4) an energy module (EM), which consists of formate dehydrogenase (FDH) from *Pseudomonas sp*. (strain 101)[60] which is involved in oxidation of formate to CO_2 and NADH. Module 1 and 2 of the RGP is considered as the core module of the pathway that is up to glycine biosynthesis. Different types of Reductive glycine pathways were designed based on possibilities of various glycine/serine assimilation routes, different energy efficiencies and different degree of overlaps with central metabolism and are presented in the recently published book chapter[61].

Nevertheless, the RGP shown in **Figure 2** is highly energy efficient and has minimal overlaps with central metabolism. This pathway is already engineered in *E. coli*, *C. necator* and established complete formatotrophic growth. A few recent studies show that this pathway is also found to be operating naturally in several specialized prokaryotes. Metagenomics analysis of enriched phosphate-oxidizing microbial community revealed that *Ca. P. anaerolimi* oxidize phosphate by donating electrons CO_2 to produced formate, which is further consumed via RGP. In this bacterium, the phosphate oxidation using CO_2 as the sole electron acceptor, coupled with native RGP might provide autotrophic growth; thus, this pathway is proposed as the seventh CO_2 fixation pathway[62]. An experimental and modeling study on acetogenic *Clostridium drakei* revealed the operation of the RGP along with the WLP (Wood-Ljungdhal pathway) while it is growing autotrophically. The GSRP (glycine synthase-reductase pathway) is involved in the synthesis of serine, acetyl-CoA, and acetyl-P[50]. Another recent multi-omics study on *Desulfovibrio desulfuricans* revealed the operation of the most efficient RGP allowing autotrophic growth of this sulfur-reducing microbe[56]. Nevertheless, all these microbes are special microbes with peculiar CO_2 reduction metabolic reactions, and these microbes are not well-established for industrial applications. Thus, engineering the RGP in well-established biotechnology hosts *S. cerevisiae* would advance the C1-based bioprocessing applications.

2.5. Yeast as a host for establishing synthetic formate assimilation

Saccharomyces cerevisiae is a biotechnologically relevant yeast. This simple eukaryote microbe is well studied for basic science and industrial applications. It is one of the most domesticated microbes and has been used by humankind for many centuries. Yeast has been utilized for decades as a prominent "metabolic cell factory" for easy cost-effective cultivation, fast and robust growth at extreme pH levels, is generally considered as safe (GRAS), and work with yeast benefits from the availability of enormous set of molecular biology tools developed for the manipulation of yeasts. Yeast is used in the production of the spectrum of value-added products, bread, chocolate, food additives, wine additives, alcohols, antioxidants and antimicrobial molecules, as comprehensively described herein[63,64]. Yeast has been exploited for its large-scale production for centuries; in World War-I for the large-scale production of glycerol: this involved the application of alkali to yeast induced fermentation of glucose to glycerol[65]. Fuel alcohols (pentanol, isobutanol, isopentanol, and several other alcohols produced in small amounts in several organisms) are produced in yeast as byproducts of amino acid metabolism[66]. They have many industrial uses, including as solvents, additives, precursors, and recently as biofuels[67]. Several studies in the past decade have been dedicated to increasing alcohol-based fuel production in yeast[68,69]. Pharmaceuticals such as human serum albumin, vaccines against hepatitis virus, and human papillomavirus[70], the antimalarial drug artemisinin[71], and hydrocortisone[72] are also produced in yeast. With the improvement in molecular biology techniques, it has become possible to produce more and increasingly complex compounds like artemisinin in yeast cells. In many cases, yeast is engineered to produce natural products through the reconstruction of biosynthetic pathways taken from other organisms. This is done by heterologous expression of the relevant pathway genes. In the early eighties, an exogenous human protein was successfully produced in yeast: the human insulin gene was expressed in yeast for the industrial production of this recombinant hormone[70,73]. Another prominent and recent example is thebaine. This opioid, routinely used as a painkiller, was successfully produced in yeast; to this end, 21 foreign genes were inserted into the yeast genome[74]. These genes encode biosynthetic enzymes originating from several species of plants, mammals, and bacteria. This work demonstrated the power and complexity of modern metabolic engineering in yeast.

Metabolic engineering is a fast-developing field. It is defined as the direct improvement of cellular properties through the modification of specific biochemical reactions or the introduction of new ones with the use of recombinant DNA technology. A plethora of compounds, many for commercial use, have been produced in microorganisms: fuels, dyes, pharmaceuticals, plastics, and other chemicals. For decades, these chemicals have been produced out of

petroleum and other fossil carbons. As *S. cerevisiae* is GRAS (Generally Considered as Safe), it can be used for food and feed applications. Thus, engineering formatotrophic growth to generate formate-dependent yeast biomass could boost the format-based circular carbon economy. So far, only module 1 (the core module) of the RGP was demonstrated in a YCG-Fr strain which is a yeast glycine biosensor strain[75]. However, a complete functional RGP is not yet demonstrated in the yeast system.

2.6. The Reductive glycine pathway in yeast cellular context

S. cerevisiae harbors all the enzymes that are involved in operating the RGP, but they operate in a different context to that required for growth on formate. In the usual conditions, when the yeast is growing on glucose, the glycolytic serine is oxidized to CH_2-THF and glycine in both cytoplasm and mitochondria. Shm1p in the mitochondrial compartment and Shm2p in the cytoplasm catalyze these reactions. Glycine is oxidized to CH_2-THF, CO_2, and NH3 by the GCV system in mitochondria while producing an NADH. Further, CH_2-THF is oxidized to formate leading to ATP and NADPH synthesis in both mitochondria, by Mis1p, and cytoplasm, by Ade3p. The resulting formate in the cytoplasm can be further oxidized to CO_2, producing NADH. The native context of this pathway is majorly operated to supply energy and reducing power for the other cellular reactions of the metabolic network.

Based on their intracellular localization, the complete RGP can be confined to mitochondria in the yeast system. Additionally, parts of the RGP also operate in the cytoplasm because of the availability of cytoplasmic isozymes (**Figure 3**). In the yeast system, the RGP is divided into four modules, (1) Glycine synthesis module, (2) Glycine to serine conversion module, (3) Serine deamination or pyruvate synthesis module, and (4) Energy module.

Module 1 can be divided into two sub-modules, i.e., the C1 maturation module and the glycine synthesis module. In the C1 maturation module, the formate is ligated to intracellular THF and further serially reduced to CH_2-THF by spending ATP and NAD(P)H. The trifunctional Mis1p carries out these reactions in mitochondria, and the trifunctional Ade3p, in the cytoplasm. This is naturally active in the yeast, which is why this module is considered as the submodule in the core module of the TGP. The CH_2-THF produced in mitochondria will be fed into the glycine synthesis module, condensed with NH_3 and CO_2 molecules to produce a glycine molecule by spending an NADH molecule. This set of reactions is carried out by Gcv1p, Gcv2p, Gcv3p, and Lpd1p enzymes, and all these enzymes are localized to mitochondria. However, Lpd1p is present in the peroxisome and other organelles of yeast as well. In the second module, the glycine produced from formate in the mitochondrial compartment can be transported to the cytoplasm. Glycine will be converted to serine by condensing with another CH_2-THF molecule

produced from formate. This reaction is carried out by Shm1p in the mitochondrial compartment and by Shm2p in the cytoplasm. In the third module, the serine available in the mitochondrial compartment will be deaminated by mitochondrial Cha1p to pyruvate, which will be used as a precursor for biomass production.

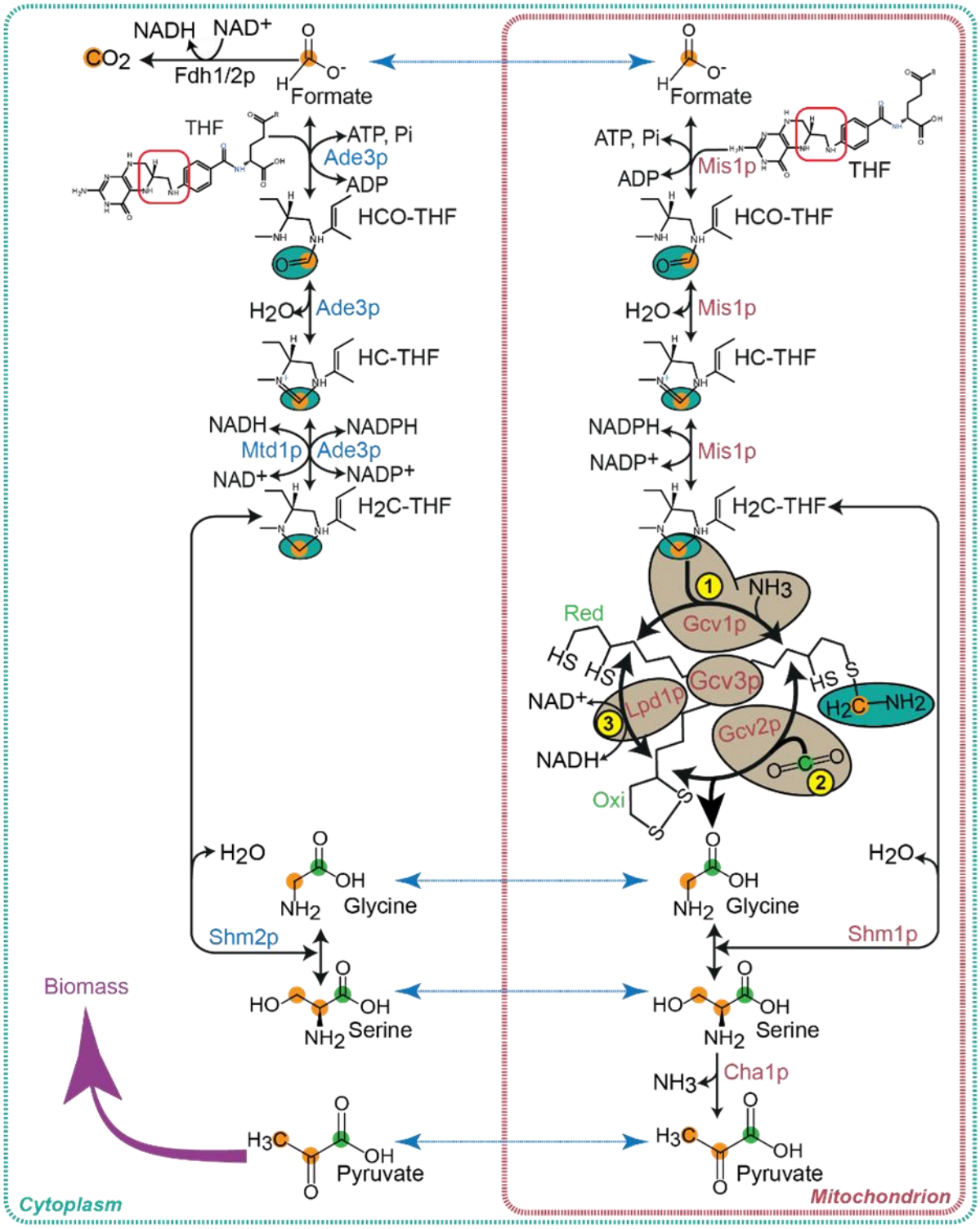


Figure 3: Distribution of the RGP modules in yeast cellular context
Formate ligation to mitochondrial THF system to produce 10, formyl-THF (CHO-THF), which is further reduced to 5, 10, methenyl-THF (CH-THF), and 5, 10, methylene-THF (CH_2-THF) in mitochondria by trifunctional **Mis1p** enzyme. The cytoplasmic trifunctional **Ade3p** enzyme carry out the same reactions in cytoplasm. Glycine Synthase System composed of **Gcv1p**, **Gcv2p**, **Gcv3p**, and **Lpd1p** involved in glycine synthesis localized to mitochondria by assimilating CH_2-THF, NH_3, and a CO_2. Glycine synthesized in mitochondria can be converted to serine by assimilating one CH_2-THF molecules via **Shm1p**. The same reaction is carried out by **Shm2p** in cytoplasm. Serine produced in mitochondria is deaminated to pyruvate by **Cha1p**.
Mis1p/Ade3p-mitochondrial/cytoplasmic trifunctional formate-tetrahydrofolate ligase/ methenyltetrahydrofolate cyclohydrolase/methylenetetrahydrofolate dehydrogenase; **Mtd1p**-cytoplasmic methylenetetrahydrofolate dehydrogenase (NADH dependent); **Gcv1p-T** subunit of the mitochondrial glycine decarboxylase complex (Aminomethyltransferase); **Gcv2p-P** subunit of the mitochondrial glycine decarboxylase complex; **Gcv3p-H** subunit of the mitochondrial glycine decarboxylase complex (Glycine dehydrogenase); **Lpd1p**-Dihydrolipoamide dehydrogenase; **Shm1p/Shm2**-mitochondrial/cytoplasmic serine hydroxymethyltransferase; **Cha1p**-serine deaminase localized to mitochondria; **Fdh1/2**-formate dehydrogenase.

The fourth module, or energy module of the RGP, oxidizes formate to CO_2 and NADH to produce and supply reducing power and energy for the operation of the RGP. Many yeast strains also harbor Fdh1/2p capable of naturally oxidizing formate in the detoxification process while gaining reducing power (**Figure 3**).

As mentioned before, the natural activity of these enzymes (Mis1p, Gcv1p, Gcv2p, Gcv3p, Lpd1, Shm1p/2p) catalyzes net flux in the reverse direction of the RGP flux that is towards the oxidation of serine and glycine to formate and CO_2, which leads to reducing power and energy synthesis in yeast.

2.6.1. Operation of Glycine Synthesis Module in the yeast system

The glycine cleavage system oxidizes glycine to CO_2, NH_3, and CH_2-THF in its natural activity, whereas reversal of this activity is required for glycine synthesis in the RGP. This is called the Glycine Synthase Complex (GSC)[61], which requires high CO_2 to produce glycine from CH_2-THF, NH_3, and CO_2.

In the initial reaction in GSC, the CH_2 group from CH_2-THF and an ammonia molecule are taken by the Gcv1p enzyme (amino methyltransferase), condensed, and attached to a reduced thiol group on the lipoate moiety of the Gcv3p, a carrier peptide of the GSC. Gcv2p (glycine dehydrogenase) assimilates a CO_2 molecule and combines it with the methylene-amine group, leading to glycine release. This leaves oxidized thiol groups on lipoate[76]. Lpd1p (oxidoreductase) reduces the thiols groups of Gcv3p to recover it for the next round of assimilations for glycine production (**Figure 3**).

3. Aim of the project

My Ph.D. project aimed to establish the Reductive Glycine Pathway in the yeast system. The starting point was the publication of Gonzalez De La Cruz et al. (2019)[75], in which the first module of the RGP, glycine biosynthesis from the formate, was tested in the *S. cerevisiae* YCG-Fr strain. Further engineering of serine biosensor strain on top of this strain was not successful due to several auxotrophies, with the consequence that a strict biosensor strain was not obtained, even when more than one strategy was employed. Thus, I procured two different prototrophic *S. cerevisiae* strains, i.e., S288c and FL100.

Progress of thesis:

1. Firstly, I constructed serine biosensor strains on both haploid yeast strains. I confirmed strict phenotype for the engineered auxotrophies, i.e., glycine and serine. I then engineered and tested module 2 of the RGP for efficient glycine to serine conversion by overexpressing three different serine hydroxymethyltransferase (SHMTs).
2. Next, I established a glycine synthesis module in combination with module 2. I conducted Adaptive Laboratory Evolution (ALE) for glycine and serine synthesis from formate. Further validated glycine, serine, and pyruvate synthesis via ^{13}C carbon tracing.
3. Then, I characterized the substrate dependency in the evolved serine biosensor strains for the RGP-dependent growth.
4. Further, the evolved strains were sent for genome sequencing to identify mutations that might potentially play a role in operating the RGP. Reverse engineering of one of the early mutations reproduced the formate-dependent growth in serine biosensor strain. I also identified the potential role of mutation on C1 assimilations.
5. I further tried to evolve the pyruvate-producing strain for complete formatotrophy, but it was not successful.
6. As a subsequent step, I engineered the third and energy modules of the RGP in the S288c strain. Then, established growth on formate and serine with short-term ALE in the S288c strain expressing serine deaminase (Cha1p) and formate dehydrogenase (Fdh1p).

4. Materials and methods

4.1. Strains used in this study

Saccharomyces cerevisiae S288c (*MATα SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6*) (procured from Dr. Patrick Yizhi Cai's lab, Manchester, UK), and FL100 (*MATa)* strain (procured from DSMZ, Braunschweig, Germany) variants were used as the base strains in this study. All other yeast strains were derived from these two parental strains and are listed in **Table 2** (Section 4.5). For cloning and plasmid propagation, *E. coli* DH5α and NEB10Beta cells (New England Biolabs GmbH, Frankfurt, Germany) were used.

4.2. Media and growth conditions

YPD medium composed of 1 % (w/v) yeast extract, 2 % (w/v) peptone, and 2 % (w/v) glucose was used for general culturing of yeast cells for DNA transformations and genome manipulations. 2x YPD (2 % yeast extract, 4 % peptone, 4 % glucose) was used for optimal growth prior to transformations. Synthetic Complete (SC) medium is composed of 1x (1.7 g/L) yeast nitrogen base (YNB) (Sigma-Aldrich, Steinheim, Germany), a customized Drop-Out mix of amino acids and nucleotides (Table S1) for optimal growth lacking the components used for selection (e.g., Glycine and Serine in SC-Gly-Ser medium), 5 g/L of ammonium sulfate ($(NH_4)_2SO_4$) as nitrogen source unless differently indicated, and appropriate carbon sources. Synthetic Minimal (SM) medium comprises 1x YNB, 5 g/L $(NH_4)_2SO_4$, and appropriate carbon sources. If not indicated otherwise, in both SC and SM media, 100 mM glucose was used as the primary carbon source, and 0.5 mM glycine, 0.5 mM serine, and 250 mM sodium formate were used as secondary carbon sources as indicated. All individual amino acids, nucleotides, and ammonium sulfate were procured from either Sigma-Aldrich or Carl Roth (Karlsruhe, Germany). All antibiotics used in yeast cultures or *E. coli* cultures were procured from Carl Roth. Yeast extract, peptone, glucose, sodium formate, ethanol, and isopropanol were procured from Carl Roth (Karlsruhe, Germany). Yeast strains were incubated at 30 ºC at 200 rpm and provided with an ambient or high CO_2 atmosphere. Geneticin (G418) (200 mg/L), hygromycin (200 mg/L), and nourseothricin (100 mg/L) were used in yeast cultures for the selection of plasmids with Kan^R, Hyg^R, and Nat^R cassettes. *E. coli* was always grown on LB medium (1% tryptone, 0.5% yeast extract, 1% NaCl) with appropriate antibiotic selection i.e., ampicillin (100 mg/L) or kanamycin (100 mg/L).

4.3. PCR and DNA preparation

PCR reactions were performed with PrimeSTAR MAX DNA Polymerase (BD Clontech GmbH, Heidelberg, Germany) or Phusion High-Fidelity polymerase (Thermo Fisher Scientific GmbH, Dreieich, Germany) following the manufacturer's recommendations. Colony PCR of *E. coli* cells was done with DreamTaq polymerase (Thermo Fisher Scientific GmbH, Dreieich, Germany). Phire Plant Direct PCR Master Mix (Thermo Fisher Scientific GmbH) was used for yeast colony PCR. All primers were synthesized by Eurofins Genomics GmbH (Ebersberg, Germany) or IDT (Leuven, Belgium) (For information on primers, refer Table S2 and Plasmid maps S1-S8).

Lucigen Master pure (Lucigen Corp., WI, USA) and YeaStar Genomic DNA Kit (Zymo research Freiburg, Germany) were used for genomic DNA isolation from yeast cells. Yeast plasmid miniprep II (Zymo Research) was used for plasmid preparation from yeast. GeneJET plasmid-miniprep-kit (Thermo Fisher Scientific GmbH) was used for plasmid preparations from *E. coli*. GeneJET PCR Purification Kit and GeneJET Gel-extraction kit (both Thermo Fisher Scientific) were used for DNA purifications.

4.4. Plasmids

The pWS series of plasmids is a gift from Prof. Tom Ellis (Imperial College, London) (Addgene numbers) (**Table 1**). pWS082 is the guide-RNA entry vector used for assembling short gRNA oligos (20 bp). pWS173-Kan^R, pWS174-Nat^R, and pWS175-Hyg^R carry the Cas9 expression cassette. These plasmids were used for sequential gene deletion of *AGX1*, *GLY1*, and *SER1* genes to construct serine biosensor strains.

The entire pCfB series of the plasmids are part of the EasyClone collection, procured from Addgene (Addgene part # 78231) and provided by Prof. Dr. Irina Borodina (Novo Nordisk Foundation Center for Biosustainability, Denmark). The pCfB2312 plasmid with Kan^R carries a Cas9 expression cassette under the *TEF1* promoter and *CYC1* terminator. The pCfB3041, pCfB3042, and pCfB3045 plasmids have sgRNA expression cassettes targeting the ChrX-3, ChrX-4, and ChrXI-3 loci, respectively, and a Nat^R selection cassette. The pCfB3034, pCfB3035, and pCfB2904 plasmids are genome integration vectors that carry ChrX-3, ChrX-4, and ChrXI-3 homology arms[77]. These plasmid backbones were amplified and used to construct genome integrative donor DNAs of the pathway modules.

The pVB05 plasmid is a genome integration vector with homology regions for the ChrX-4 locus made from the pCfB3035 backbone. The *SHM1* overexpression module, under *PGK1* promoter and terminator, was amplified using pLH52 as a template and assembled between

ChrX-4 homology regions in the pCfB3035 plasmid backbone to construct the pVB05 plasmid (**Map S1**). Similarly, the *SHM2* overexpression module, under *GMP1* promoter and terminator, was amplified from the pLH53 template and cloned in pCfB2904 between the ChrXI-3 homology regions to create pVB06 plasmid (**Map S2**). To construct the pVB07 plasmid (**Map S3**), the *EcGlyA* expression cassette under the control of the *RPL3* promoter and *CWP2* terminators was amplified from the pVB01 plasmid (**Map S4**) and assembled in the pCfB3035 plasmid between the homology regions for genome integration into ChrX-4 locus. These plasmids were constructed using the NEBuilder HiFi DNA assembly master mix as per the supplier's recommendation (New England Biolabs GmbH, Frankfurt, Germany). The pLH52 and pLH53 plasmids were a gift from Dr. Lena Hochrein (University of Potsdam, Germany).

The pFM340 plasmid (**Map S5**) carries expression cassettes *MIS1*, *GCV1*, *GCV2*, and *GCV3* of the glycine synthesis module of the RGP. The pFM340 plasmid was created by replacing the *URA3* selection marker on the original pJGC3 plasmid (**Map S6**) with a Hyg^R cassette. The pJGC3 plasmid was initially constructed using the AssemblX methodology as optimized in Hochrein et al.[78] and tested successfully for glycine synthesis[75]. The plasmid region of pFM340, encoding *MIS1*, *GCV1*, *GCV2*, and *GCV3,* was amplified and cloned in the pCfB3034 between the ChrX-3 homology sequences to construct the pFM275 plasmid (**Map S7**) using NEBuilder HiFi DNA assembly methodology.

The *CHA1* expression (Serine deamination module) under the control of *TDH3* promoter and *CYC1* terminator and *mtCbFDH1* expression (formate dehydrogenase - energy module) under the control of *PGK1* promoter and *ADH1* terminator were assembled in the pVB12 plasmid (**Map S8**). *CHA1*, TDH3p, PGK1p, CYC1t, and ADH1t were amplified from the genome of *S. cerevisiae* or the standard yeast vectors carrying these promoter and terminator sequences. The formate dehydrogenase gene sequence of methylotrophic yeast *Candida boidinii* was extracted from NCBI. This sequence was codon-optimized for *S. cerevisiae* and synthesized from TWIST bioscience (California, USA). The *CbFDH1* gene was amplified by fusing with the *PreCoxIV* mitochondrial leader sequence (mt) to target Fdh1p activity in the mitochondria and assembled in the pVB12 plasmid (**Map S8**). Fluorescence microscopy analysis suggests that the *GFP* fused with the *PreCoxIV* leader sequence at the amino-terminus end was localized to mitochondria (**Figure S 1**). The *PreCoxIV* leader sequence was tested as the responsible signal sequence for importing the Cytochrome c oxidase subunit IV to mitochondria. This subsequently is cleaved in the mitochondrial matrix upon import[79].

4.4.1. Yeast transformation

DNA transformations into yeast were performed using the lithium acetate-based heat shock method[80]. Recovered transformants were selected based on the selection marker available, and colonies were screened for mutations by PCR amplification and Sanger sequencing of the amplified locus. Important primer used for screening and sequencing are listed in the Supplementary information (**Table 2**).

Table 1: Plasmids used in this study
List of plasmids used for gene deletions, construction of pathway modules and genomic integration of the pathway modules into specified chromosomal location. *ox*-overexpression; Hr-homology regions; Chr-chromosome; mt-mitochondria targeting leader sequence; **Cb**-*Candida boidinii*. **Ec**-*E. coli*;

Table 1 Plasmids used in this study			
Plasmid	**Details**	**Use**	**Source**
pWS082	sg entry vector	Gene deletion	Addgene # 90516
pWS173	Cas9-Kan^R	Gene deletion	Addgene # 90960
pWS174	Cas9-Nat^R	Gene deletion	Addgene # 90961
pWS175	Cas9-Hyg^R	Gene deletion	Addgene # 90962
pCfB2312	Cas9-Kan^R	Pathway integration	Addgene Kit # 1000000098
pCfB3041	Nat^R-sgChrX-3	Target-ChrX-3	Addgene Kit # 1000000098
pCfB3042	Nat^R-sgChrX-4	Target-ChrX-4	Addgene Kit # 1000000098
pCfB3045	Nat^R-sgChrXI-3	Target-ChrXI-3	Addgene Kit # 1000000098
pCfB3034	HrChrX-3	ChrX-3 donor	Addgene Kit # 1000000098
pCfB3035	HrChrX-4	ChrX-4 donor	Addgene Kit # 1000000098
pCfB2904	HrChrXI-3	ChrXI-3 donor	Addgene Kit # 1000000098
pFM275	Hyg^R-MIS1-GCV1-3ox	Pathway expression	In this study
pVB05	PGK1p-*SHM1ox*	SHM1-ChrX-4 donor	In this study
pVB06	GMP1p-*SHM2ox*	SHM2-ChrXI-3 donor	In this study
pVB07	RPL3p-*EcGlyAox*	EcGlyA-ChrX-4 donor	In this study
pFM340	Hyg^R-*MIS1-GCV1-3ox*	Pathway expression	In this study
pVB12	*CHA1-mtCbFDH1*-Kan^R	Pathway expression	In this study

4.5. Strain engineering

The serine biosensor (ΔS) strain, VBS01 (S288c_ΔS) (**Table 2**), was generated in the *S. cerevisiae* S288c background by making knockouts of *AGX1*, *GLY1*, and *SER1* genes to block

the serine and glycine routes. All knockouts were verified by PCR and Sanger sequencing (at LGC Genomics, Berlin, Germany). VBS01 was transformed with pFM340, encoding for the glycine synthesis module of the RGP, to make the VBS02 strain. To engineer further modules of the RGP in the serine biosensor strain, pCfB2312 was transformed into VBS01 to create VBS03. The VBS04 strain was constructed by genome integration of the *SHM1* module into the ChrX-4 locus of the VBS03 strain. The *SHM1* module, with ChrX-4 homology arms, was amplified from pVB05 and transformed into VBS03 along with the pCfB3042 plasmid. Transformants were selected based on G418, and nourseothricin resistances were screened for the successful integration events by colony PCR and confirmed by Sanger sequencing.

VBS05 strain was constructed by integrating the *SHM2* module into the ChrXI-3 locus of VBS03. The *SHM2* overexpression module, with ChrXI-3 homology regions, was amplified from pVB06 and transformed into VBS03 along with the pCfB3045 plasmid. The *EcGlyA* module, with Chrx-4 homology arms, was amplified from pVB07 and transformed into VBS03 along with the pCfB3042 plasmid. All three strains, VBS04, VBS05, and VBS06, overexpress one of the three SHMT modules (*SHM1*/*SHM2*/*EcGlyA*) involved in the glycine to serine conversion module of the RGP. Similarly, the VBS07 strain was constructed by genome integration of glycine synthesis module amplified from pFM275 plasmid Chx-3 homology regions. To further construct serine biosensor strains expressing glycine and serine synthesis modules, the VBS08, VBS09, and VBS10 strains were made by transforming the pFM340 plasmid into VBS04, VBS05, and VBS06 strains, respectively. VBS11 carried *URA3* knockout on top of the S288c_ΔS genotype, further transformed with pFM345-URA3 to make the VBS12 strain. Additionally, VBS17 and VBS19 strains were created by making a knockout of 108 bp from the CDS of the *GDH1* gene on genomes of VBS06 and S288c_WT strain, respectively, using the pWS plasmid system. Then, VBS17 is transformed with pFM340-Hyg^R to make VBS18.

The S288c_WT was transformed with pVB12 plasmid expressing *CHA1* and *mtCbFDH1* to construct the VBS20 strain to engineer the serine deamination (pyruvate synthesis) and energy modules.

Two serine biosensor strains were constructed on FL100_WT in a slightly different approach. The VBS13 strain was created by knocking out *SER1*, *AGX1*, *GLY1*, and *SHM1* in the FL100 background. VBS14 carried *SHM2* knocked instead of *SHM1* knockout and other common mutations essential to creating a ΔS strain. Both variants of FL100_ΔS were transformed with pFM340 to create VBS15 and VBS16 strains to express the glycine synthesis module.

Table 2: List of yeast strains constructed for this study

Δ-deletion; ΔS-serine biosensor strain (cannot produce serine glycine from glucose). SHMT-Serine hydroxymethyltransferase; Ec-*E. coli*;

Table 2: List of yeast strains used in this study Name	Strains	Genotype	Glycine synthesis module overexpression	Plasmid	SHMT overexpression module	Chromosomal integration
VBS01	S288c_ΔS	S288c *Δagx1, Δgly1, Δser1*	No	—	No	—
VBS02	S288c_ΔS-pFM340-Hyg^R	S288c *Δagx1, Δgly1, Δser1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	No	—
VBS03	S288c_ΔS-pCfB-Cas9-Kan^R	S288c *Δagx1, Δgly1, Δser1*	No	pCfB2312	No	pCfb2312
VBS04	S288c_ΔS-ChrX-4::SHM1	S288c *Δagx1, Δgly1, Δser1*	No	—	*SHM1*	ChrX-4
VBS05	S288c_ΔS-ChrXI-3::SHM2	S288c *Δagx1, Δgly1, Δser1*	No	—	*SHM2*	ChrXI-3
VBS06	S288c_ΔS-ChrX-4::EcGlyA	S288c *Δagx1, Δgly1, Δser1*	No	—	*EcSHMT* (*GlyA*)	ChrX-4
VBS07	S288c_ΔS-ChrX-3::GCV1-3-MIS1	S288c *Δagx1, Δgly1, Δser1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	ChrX-3	No	ChrX-3
VBS08	S288c_ΔS-ChrX-4::SHM1-pFM340-Hyg^R	S288c *Δagx1, Δgly1, Δser1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	*SHM1*	ChrX-4
VBS09	S288c_ΔS-ChrXI-3::SHM2-pFM340-Hyg^R	S288c *Δagx1, Δgly1, Δser1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	*SHM2*	ChrXI-3
VBS10	S288c_ΔS-ChrX-4::EcGlyA-pFM340-Hyg^R	S288c *Δagx1, Δgly1, Δser1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	*EcSHMT* (GlyA)	ChrX-4
VBS11	S288c_ΔSU	S288c *Δagx1, Δgly1, Δser1, Δura3*	No	—	No	—
VBS12	S288c_ΔSU-pFM345-Ura3	S288c *Δagx1, Δgly1, Δser1, Δura3*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM345	No	pFM345
VBS13	FL100_ΔS.1	FL100 *Δagx1, Δgly1, Δser1, Δshm1*	No	—	No	—
VBS14	FL100_ΔS.2	FL100 *Δagx1, Δgly1, Δser1, Δshm2*	No	—	No	—
VBS15	FL100_ΔS.1-pFM340	FL100 *Δagx1, Δgly1, Δser1, Δshm1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	No	—
VBS16	FL100_ΔS.2-pFM340	FL100 *Δagx1, Δgly1, Δser1, Δshm2*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	No	—
VBS17	VBS06_Δgdh1::108bp	*S288c Δagx1, Δgly1, Δser1, Δgdh1*	No	—	No	—
VBS18	VBS10_Δgdh1::108bp	S288c *Δagx1, Δgly1, Δser1, Δgdh1*	*GCV1* , *GCV2* , *GCV3* , *MIS1*	pFM340	*EcSHMT* (GlyA)	ChrX-4
VBS19	S288c_WT_Δgdh1::108bp	S288c *Δgdh1*	No	—	No	—
VBS20	S288c_WT-pCHA1ox-mtCbFDH1ox-Kan^R	S288c WT	No	pVB12	No	pVB12

4.5.1. CRISPR-Cas9 methodology for gene deletions

For gene knockout using CRISPR-Cas9 methodology, pWS082, pWS173, pWS174, and pWS175 system was used. This CRISPR-Cas9 methodology is optimized by combining principles and tools described in[81–83]; (https://benchling.com/pub/ellis-crispr-tools). *SER1*, *AGX1*, and *GLY1* genes were deleted by making marker-free complete CDS knockouts.

The sgRNA oligos (**Table S2**) were designed using Yeastriction online tools [84] (http://yeastriction.tnw.tudelft.nl/#!/) and Benchling CRISPR tool for the targeted genes (http://www.benchling.com), e.g., *SER1*, *AGX1*, *GLY1*, *SHM1*, *SHM2*, and *GDH1*. The GACTTT sequence is added to the upper oligo's 5' end of the upper oligonucleotides. CAAA is added to the 5' end and AA to the 3' end of the lower primer to complete the HDV ribosomal sequence for stability sgRNA[83,85]. Then both primers were annealed in a thermocycler by a stepwise decrease of temperature from 95 ºC to 25 ºC.

```
1    5' GACTTTNNNNNNNNNNNNNNNNNNN 3'
                |||||||||||||||||||
2          3' AANNNNNNNNNNNNNNNNNNNCAAA 5'
```

The annealed double standard guide oligonucleotides with GACT and CAAA overhangs are assembled into the pWS082 vector. A 100 ng of pWS082, 2 µL of annealed oligos with overhangs, 1 µL of T4 DNA ligase, and *Esp*3I restriction enzyme were mixed in 1x T4 ligase buffer. The combined restriction/ligation was performed in a thermocycler: 30 x (5min at 30°C, 5 min at 16°C)[81]. Subsequently, this Golden-Gate reaction was transformed into competent *E. coli* DH5α and selected for ampicillin-resistant transformants. Non-GFP colonies were picked as positive clones considering the GFP expression cassette is replaced by sgRNA oligonucleotide insertion. The plasmid was prepared from these non-GFP colonies, and oligonucleotide insertion was confirmed by sequencing. The verified pWS082 plasmids with gDNA inserts were used as templates to amplify the whole sgRNA expression cassettes for transformation.

The Cas9 expression plasmids, i.e., pWS173 with Kan^R, pWS174 with Nat^R, and pWS175 with Hyg^R, carry a 2µ *Ori*, a nuclear targeted Cas9 expression cassette, and a super-folder GFP expression cassette cloned between *Esp*3I restriction sites. To prepare Cas9 backbones with different selection markers, these plasmids were digested with *Esp*3I, and the backbone DNA fragment containing the Cas9 cassette and the respective selection marker were gel purified. This fragment carries homology sequences, at both ends, to the amplified sgRNA expression cassettes from pWS082 vectors.

To delete a targeted gene or partial DNA sequence from the genome, repair DNA fragments were designed using Yeastriction online tool, SnapGene 6.02 (from Insightful Science; available at snapgene.com)., or Geneiouse 11.1.5 software (https://www.geneious.com). A 60 bp of the upstream region and a 60 bp of the downstream region of the targeted deletion were combined separately from both DNA strands and generated a pair of 120 bp long primers complementing each other. The primers were annealed and used as double-stranded donor DNA for genome repair. If necessary, point mutations were introduced to destroy PAM sequences and prevent the potential Cas9 binding to the already edited locus or the free double-stranded donor DNA fragments.

The linear sgRNA fragment, linear Cas9 backbone, and linear donor/repair DNA fragments were transformed in the intended yeast strain and selected according to the resistant cassettes. Due to the mutual homology sequences, the sgRNA fragment and Cas9 fragment form a single circular plasmid by activating the native recombination machinery of yeast[86]. These transformants can be selected based on the antibiotic resistance cassette as the circular plasmid express resistance marker, Cas9 protein, and sgRNA. The tRNA fragment undergoes self-cleavage upon expression, leaving a mature, stable, persistent sgRNA on which Cas9 is assembled. The sgRNA and Cas9 complex introduces a double standard break in the genome, which are repaired by already activated recombination machinery according to the homology regions on the transformed donor DNA.

4.6. Adaptive Laboratory Evolution

Adaptive Laboratory Evolution (ALE) was conducted with VBS08, VBS09, VBS10, and several other strains in the SC-Ser-Gly medium. SC medium used for ALE is composed of 1x YNB, 100 mM $(NH_4)_2SO_4$, 100 mM glucose, and 100 or 250 mM formate. SC medium contains a Drop-Out mix of amino acids, including standard concentrations of Uracil or Adenine, in defined quantities (Sigma-Aldrich, 2014-Y1751 & Y1856) (**Table S1**). A custom mix of Drop-Out without serine and glycine was prepared and used in the SC-Ser-Gly medium to establish glycine and serine synthesis from formate using growth as a readout. The maintenance of the glycine synthesis module of the RGP encoded on pFM340 was achieved with hygromycin selection (200 mg/L).

4.6.1. Determination of growth improvements in the ALE

The growth improvements of VBS10 strains, operating the RGP for the formate-dependent synthesis of serine and glycine were assessed through high throughput growth experiments. The growth rates of VBS10 strains from Ev03, Ev06, Ev12, Ev16, and Ev21 were evaluated by screening multiple colonies from each selected evolution round using multi-well readers.

The glycerol stocks of the selected evolution rounds were streaked on YPD hygromycin plates and incubated at 30 ºC until prominent colonies appeared. 10 colonies from each selected evolution round were picked and inoculated in SM medium with 50 mM glucose and 250 mM formate for primary culture preparation. These cultures were incubated under a 10% CO_2 atmosphere at 30 ºC with 200 rpm of orbital shaking until growth appeared. The precultures were washed in sterile distilled water and used as seed culture to determine the growth rate in SM medium with 50 mM glucose with or without 100 mM sodium formate. All the 10 cultures were inoculated to 0.02 OD_{600} in a 1.5-2 ml Eppendorf tubes in the selective medium and mixed. The homogenous cultures were distributed in 48 well plates in technical duplicates. To prevent evaporation during the experiment, each well was filled with a culture volume of 350 µl and overlayered with 250 µl of mineral oil. The plates were incubated in the multi-well plate reader instruments (Tecan – Spark/infinite, Crailsheim, Germany or BioTek Epoch 2 - BioTek, Bad Friedrichshall, Germany) at 30 ºC, providing a 10% CO_2 atmosphere. Alternating orbital and linear shaking (~300 rpm, 60 sec per shaking mode) was programmed, and growth was measured at ~15 mins intervals by reading optical density at 600 nm wavelength in each cycle. The output data of the optical density values were used to plot growth curves using a custom MATLAB script (available on request) (MATLABR2019a). In the MATLAB script, the OD values of the 48 well plates were extrapolated to cuvette OD values using a predetermined conversion factor. The MATLAB script also calculates the growth rate and doubling time. The growth rates of multiple colonies from selected evolution rounds were plotted and analyzed by one-way ANOVA followed by Tukey multiple comparisons tests using GraphPad Prism version 9 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com).

4.7. Stable ^{13}C isotope tracing of formate and CO_2 utilization

To confirm the metabolic pathway responsible for the synthesis of serine and glycine in the evolved VBS10 strains, stable ^{13}C isotope labeling experiments were conducted with the evolved VBS10 strains from different evolution rounds. The VBS10 strains from Ev02, Ev06, Ev10, Ev16, Ev21, Ev25, and Ev29 were cultured in the SM medium supplied with ^{13}C or ^{12}C formate, with 40 mM glucose as the primary carbon source. The cultures were incubated in a 10% $^{13}CO_2$ or $^{12}CO_2$ atmosphere. A vacuum desiccator (Lab Companion, MA, USA) was used to grow cultures in $^{13}CO_2$, where the original gas was drawn-out by a vacuum pump, followed by refilling with 10% $^{13}CO_2$ and 90% air. Cells equivalent to 2 OD_{600} from grown cultures were collected from the early stationary phase and centrifuged at 20,000 x g for 5 mins to pool the cell biomass. The washed biomass was dissolved in 1 ml 6N HCL and hydrolyzed overnight at 95 °C. Then samples were dried at the same temperature with or without flushing air [88]. After drying the samples at 95 ˚C, the samples were resuspended in ultra-pure water, filtered,

and the masses of amino acids were analyzed with UPLC-ESI-MS as previously described[89]. Hydrolyzed proteinogenic amino acids were separated using the Waters Acquity UPLC system (Waters, Milford, MA, USA), equipped with an HSS T3 C18 reversed-phase column (100 x 2.1 mm2, 1.8 mm; Waters, Eschborn, Germany). In the mobile phase, 0.1% formic acid in H_2O (A) and 0.1% formic acid in acetonitrile (B) were used. The flow rate was 0.4 mL/min, and the gradient was: 0 to 1 min – 99% A; 1 to 5 min – linear gradient from 99% A to 82%; 5 to 6 min – linear gradient from 82% A to 1% A; 6 to 8 min – kept at 1% A; 8 to 8.5 min – linear gradient to 99% A; 8.5 to 11 min – re-equilibrate. Mass spectra were acquired using an Exactive mass spectrometer (ThermoScientific, Dreieich, Germany) in positive ionization mode, with a scanning range of 50.0 to 300.0 m/z. Spectra were recorded during the first 5 min of the LC gradients. Under these conditions, the retention times for amino acids were determined by analyzing amino-acid standards (Sigma-Aldrich) under the same conditions. Data analysis was performed using Xcalibur (ThermoScientific, Dreieich, Germany).

4.8. Genome sequencing analysis

Genomic DNA samples (isolated as in section 4.3) were run through an additional RNase treatment to clean the contaminant RNA before sending for whole-genome sequencing. For this, samples resuspended in TE buffer and RNase were added and incubated at 37ºC for one hour[90]. Then samples were precipitated and resuspended in TE buffer to send for NGS. Library construction of genomic DNA samples and sequencing was performed by Novogene (Cambridge, United Kingdom) using the paired-end Illumina sequencing platform. The sequencing data was analyzed using the breseq pipeline[91] and modified for multi chromosomal haploid organisms. The S288c genome sequences downloaded from the NCBI database were used as the reference genome to compare the genome changes acquired in the ALE. A non-modified wild-type strain was sequenced along with the evolved and unevolved strains as a control. NGS data of WT, unevolved VBS10, and evolved VBS10 strains were aligned to the S288c reference sequence and compared mutations. All mutations that occur 100% reads in all four colonies (four biological replicates) from each evolution round were considered mutations that might support formate-dependent growth improvement directly or indirectly. The list of mutations identified was segregated based on their location, i.e., either on the coding sequences or in the intergenic regions of genomes.

5. Results

5.1. Establishing the RGP in *S. cerevisiae* S288c

The *S. cerevisiae* strain is a widely used yeast for fundamental research and for developing industrial applications. The S288c variant is one of the most studied yeast from which multiple laboratory strains were derived. This strain is also applied for metabolic engineering and synthetic biology research. The S288c strain is also used for the synthetic yeast project to construct synthetic chromosomes and also a well-annotated complete genome sequence is available for this strain[92–94]. Here, I employed the prototrophic S288c strain to establish synthetic formate assimilation by expressing the RGP. Aiming to establish the formatotrophic growth in the S288c strain, I constructed different serine biosensor strains and established a selection base formate assimilation which is elaborated in the following sections.

5.1.1. Construction of a serine biosensor strain in S288c

S. cerevisiae has three different serine and glycine biosynthesis routes: Serine biosynthesis via 3-PGA is the main source of serine during growth on glucose as the main carbon source and it involves Ser1p, Ser2p and Ser33p/Ser3p (**Figure 4-I**)[95,96]. Two alternative serine synthesis pathways operate via glycine, where glycine is either produced by threonine cleavage and/or by transamination of glyoxylate[97,98]. The threonine cleavage pathway is a major glycine biosynthesis pathway when yeast is growing on glucose and is carried out in the cytoplasm by threonine aldolase (Gly1p) resulting in glycine and acetaldehyde synthesis (**Figure 4-II**). In the third pathway, glyoxylate that is produced from isocitrate by isocitrate lyase (Icl1p) is transaminated to glycine via *alanine:glyoxylate* aminotransferase (Agx1p) in mitochondria (**Figure 4-III**)[99]. This glycine biosynthesis pathway is inhibited in yeast cells exponentially growing on glucose but is the major glycine biosynthesis route during growth on ethanol or acetate. Glycine produced via threonine cleavage or glyoxylate transamination is converted to serine by Shm1p in mitochondria and Shm2p in cytoplasm[100–106]. To construct a serine biosensor strain, *SER1*, *GLY1*, and *AGX1* gene knockouts were made in the genome of *S. cerevisiae* S288c strain to disrupt serine and glycine biosynthesis via all three pathways. Complete CDS deletion of these genes were made by using a CRISPR-Cas9 system.

The combination of all three knockouts is expected to render the strain auxotrophic for serine, glycine and C1 units due to the insulation of these metabolites from the central metabolic network (Figure. 4 highlighted in the box). Hence, S288c *agx1Δ gly1Δ ser1Δ* (VBS01) strain (**Table 2**) should not be able to grow on glucose as the sole carbon source because the glycolytic fluxes are blocked from entering in to the serine, glycine and C1 metabolic network.

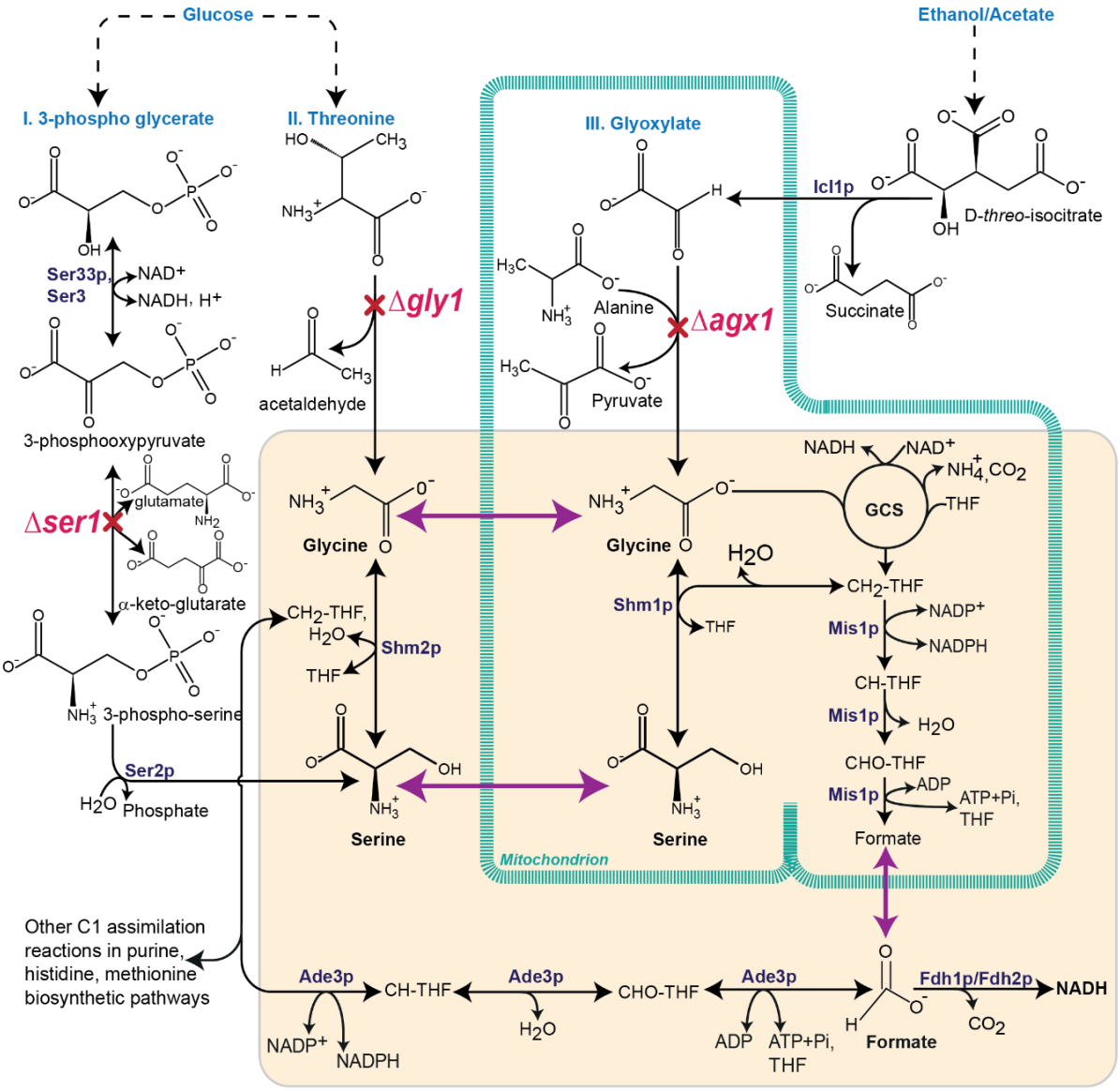


Figure 4: **Metabolic scheme of serine biosensor strain**

Three different major serine biosynthesis routes are depicted; **I.** serine biosynthesis from 3-phospho-glycerate via 3-phosphooxypyruvate and 3-phospho-serine intermediates from glucose. **II.** Serine biosynthesis from threonine via glycine intermediate, a secondary glycine and serine synthesis route when yeast is growing on glucose. **III.** Mitochondrial serine biosynthesis route from glyoxylate via glycine; this is a glycine and serine biosynthesis route when yeast is growing on ethanol or acetate as a major carbon source. The violet colour indicates the native enzymes involved in the serine biosynthesis pathways. Purple colour arrows indicate the inter-compartmental transport of the glycine, serine, and formate. Mutations in serine biosensor strain are indicated with red-cross disrupting serine or glycine synthesis from the respective pathway. Highlighted in the yellow shadow is the insulated metabolism from the central metabolic network upon the *ser1*, *gly1*, and *agx1* knockout indicated with the red crosses.

In this strain, as serine and glycine cannot be produced from glucose, intracellular C1 units are also not available, since they are produced by glycine and serine degradation. This affects the biosynthesis of cysteine, tryptophan, and phospholipids where serine is assimilated, as well as the purine, thymidylate, methionine, and histidine biosynthesis pathways, where C1 units are assimilated. As a result, this strain cannot grow on glucose as the sole carbon source but requires the supplementation of serine or glycine as a secondary carbon source.

5.1.2. Validation of serine biosensor strain

To confirm the expected metabolic phenotype of VBS01 genotype, a growth experiment was performed employing the VBS01 and S288c WT strains in SM medium (with 100 mM glucose as carbon source and 40 mM $(NH)_2SO_4$ as nitrogen source) in ambient CO_2 condition. As expected, the WT strain grows on glucose as the sole carbon source, whereas VBS01 does not grow in SM medium. Addition of glycine or serine as a secondary carbon source enabled growth of the VBS01 (**Figure 5B**), confirming the expected auxotrophy: When the SM medium is supplemented with serine it is not only assimilated directly for biomass, but can also be used for glycine synthesis through Shm1p/Shm2p (SHMTs) activity in (**Figure 5A**) mitochondria and cytosol, thereby releasing C1-units (CH_2-THFs) which can be used for purine, methionine, and histidine biosynthesis. Glycine produced by SHMTs from serine or directly supplemented in the medium, can be assimilated for biomass and can also be degraded into CO_2, NH_3, and CH_2-THF in mitochondria by the glycine decarboxylase system (GDS). The mitochondrial pool of CH_2-THF can be converted into formate by Mis1p and transported to the cytosol where it can be utilised by Ade3p to produce a cytosolic CH_2-THF pool (**Figure 5A**). Whereas Shm1p and/or Shm2p can utilize the CH_2-THF pools and convert glycine to serine in mitochondria and in the cytoplasm (**Figure 5A**). As the VBS01 strain cannot grow without glycine or serine supplementation, it can be employed for testing novel synthetic metabolic pathways for serine and glycine biosynthesis using growth as a readout. The strain can also be used for evolving novel enzymes or new to nature pathways which can be used for establishing, serine, glycine and C1 unit biosynthesis pathways.

In accordance with above observations, this strain is henceforth designated as base serine biosensor strain (ΔS). In this scientific work, I applied S288c_ΔS (VBS01) to establish the glycine and serine biosynthesis from formate and CO_2 via the RGP.

Operation of the RGP requires the presence of formate as carbon source and a high CO_2 atmosphere for reversal of the native GCV/GDC system in order to assimilate CO_2 for glycine synthesis. To further investigate the VBS01 strain in the conditions similar to the anticipated growth conditions required for the RGP activity, another growth experiment was conducted in

the presence of 250 mM formate under 10% CO_2 atmosphere in SM medium supplemented with serine or glycine. The VBS01 strain did not grow without glycine or serine supplementation, but growth was rescued when 0.5 mM serine was provided to the medium (**Figure 5C**).

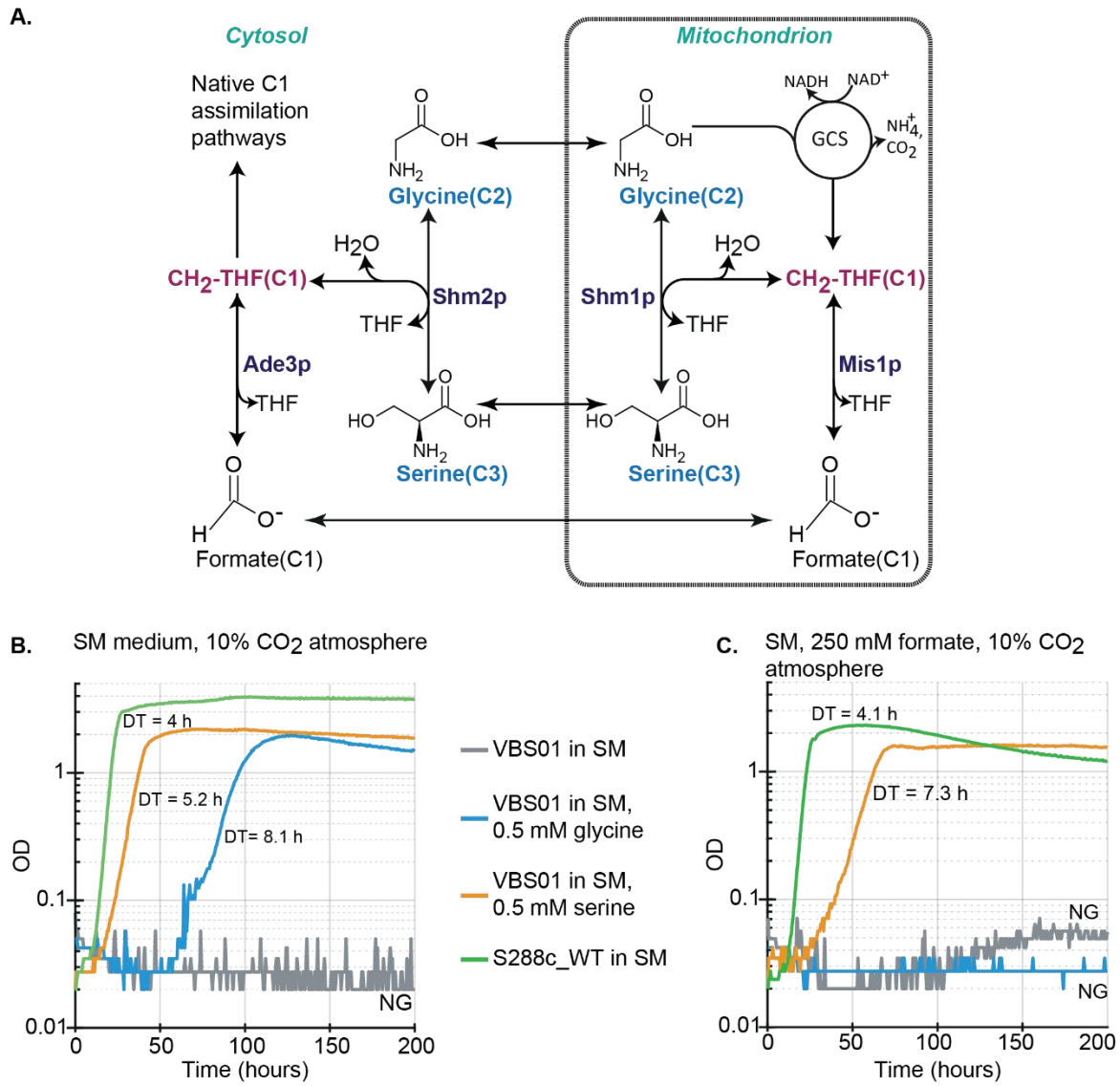


Figure 5: Metabolic scheme of glycine/serine in VBS01 strains and metabolic phenotype

A. Metabolic scheme of possible flux distribution of serine for glycine and C1 synthesis and the flux distribution of glycine for serine and C1 units to satisfy the serine, glycine, and C1 auxotrophies in VBS01 (S288c_ΔS) strains. The blue colour indicates glycine or serine as additional carbon sources. **B.** Growth curves of VBS01 strain in SM medium supplemented with glycine or serine in a high CO_2 atmosphere along with WT control without supplementation of serine or glycine. **C.** Growth curves of VBS01 strain in SM medium with formate and glycine or serine. Growth curve of WT control in SM medium with formate without glycine supplementation in a high CO_2 atmosphere. **DT** - Doubling Time of the respective strain in the respective growth condition; **NG** - No Growth; **SM** - synthetic minimal medium; **SM** medium always contains 1xYNB and 100 mM glucose unless stated otherwise; **YNB** - Yeast Nitrogen Base contains ~40 mM ammonium sulphate ($(NH_4)_2SO_4$) unless stated differently.

Surprisingly, growth was not rescued when 0.5 mM glycine is supplemented in presence of formate (**Figure 5C**), in contrast to the previous observation where 0.5 mM glycine addition to the medium without formate rescued growth of VBS01 strain (**Figure 5B**). This result indicates that serine is maybe not produced from glycine when formate is present in the medium thus growth is not supported. I speculate that the formate addition to the medium led to high CH_2-THF, which might have inhibited *SHM2* expression; however, this is not investigated further. Nevertheless, glycine conversion to serine in presence of formate is essential for serine production from formate via the RGP. Hence, it was necessary to engineer a serine biosensor strain that can convert glycine to serine efficiently in presence of formate and high CO_2 atmosphere. Thus, I proceeded to engineer the glycine to serine conversion module in the VBS01 strain by overexpressing different serine-hydroxymethyltransferases (*SHMTs*) from yeast and *E. coli* to establish efficient glycine to serine conversion in presence of formate and high CO_2.

5.1.3. Engineering SHMT modules in the VBS01 strain

To engineer an active serine synthesis module for establishing glycine conversion to serine by the assimilation of CH_2-THF synthesized from formate with the supplied glycine, I decided to overexpress different serine hydroxy methyltransferases (*SHMTs*) i.e., *SHM1* and *SHM2* from yeast and *EcGlyA* – the SHMT from *E. coli*. *ScSHM1*, *ScSHM2*, and *EcGlyA* genes were cloned in genome integrative vectors under control of the strong, constitutive promoters PGK1p, GPM1p, and RPL3p, respectively. The resulting overexpression modules of *ScSHM1*, *ScSHM2*, and *EcGlyA* were integrated into the genome of the VBS01 strain using the CRISPR-Cas9 methodology as described in methods section (**4.5.**). This resulted in serine biosensor strains with different SHMT overexpression modules (S288c_ΔS-SHMTox) that were named as VBS04 (S288c_ΔS–SHM1), VBS05 (S288c_ΔS–SHM2), and VBS06 (S288c_ΔS–EcGlyA) (**Table 2**). To test the glycine to serine conversion in S288c_ΔS-SHMTox strains, a growth experiment was conducted in SM medium with VBS01 as a control without *SHMTox*. In total four different ΔS strains were employed i.e., VBS04, VBS05, VBS06, and VBS01 and tested for conversion of glycine to serine under 10% CO_2 atmosphere with and without glycine and/or formate supplementation using growth as readout.

All the ΔS strains grew well when 0.5-1 mM glycine was added to the medium irrespective of the SHMTs over-expression (**Figure 6B**). Glycine addition to medium without formate rescued growth by satisfying glycine, C1, and serine requirements as shown in the metabolic scheme (**Figure 6A**). The VBS06 strain was growing faster at 3.5 h of doubling time as compared to all other ΔS strains i.e., VBS04, VBS05, and VBS01 which grew at 4.8 h, 4.5 h, and 7.7 h of doubling times, respectively. As expected, no growth is observed from any of the four ΔS

strains when glycine was not provided in the medium (**Figure 6B**).

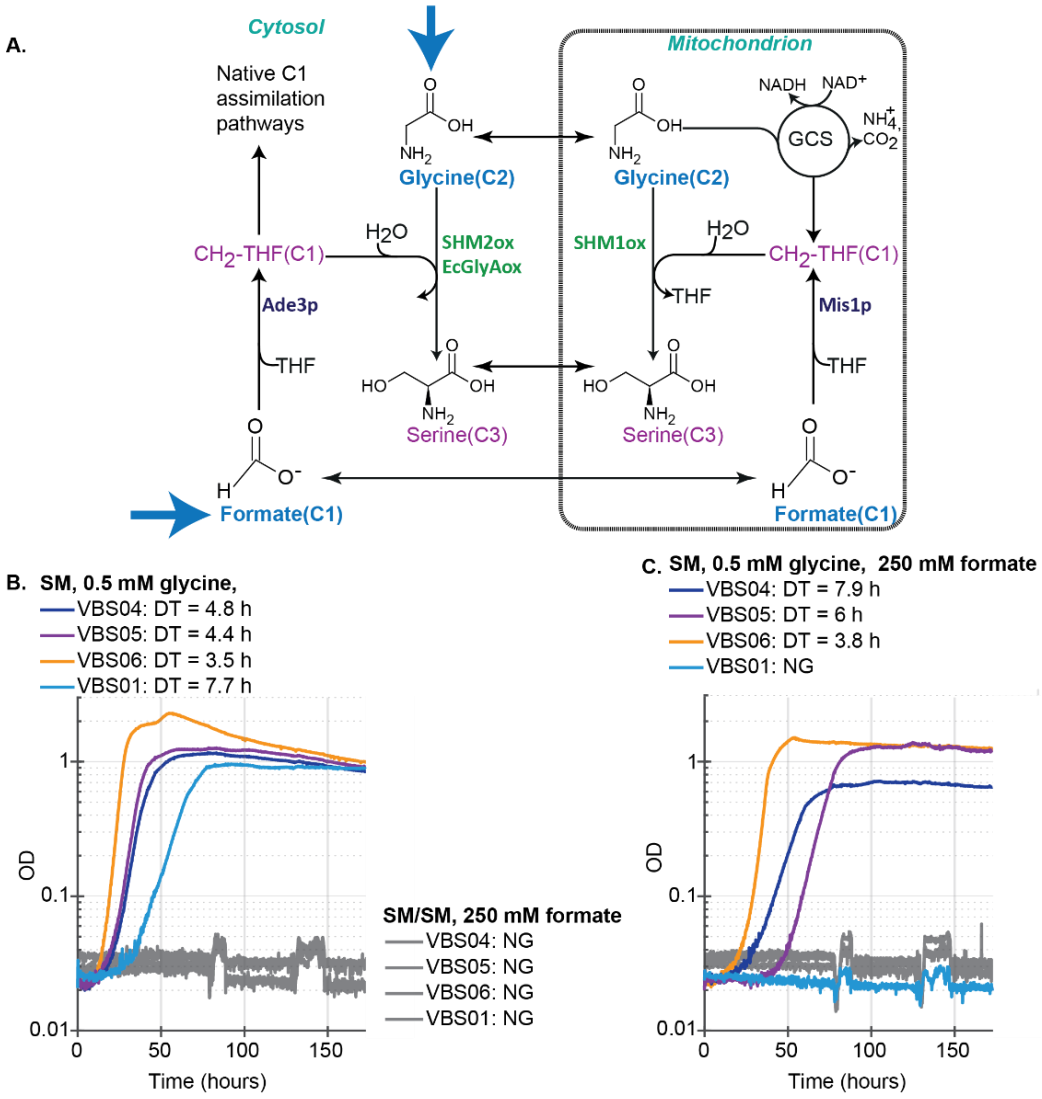


Figure 6: **Comparison of glycine conversion in SHMT overexpressing serine biosensor strains**
A. scheme showing flux distribution of glycine and formate for serine biosynthesis to satisfy the glycine, serine, and C1 auxotrophies in S288c_ΔS-SHMTox strains. The blue color indicates glycine and formate as additional carbon sources. The purple color indicates serine and C1 units to be produced from additional carbon sources. **B.** Growth curves of three different serine biosensor strains overexpressing SHMTs (i.e., VBS04, VBS05, and VBS06) along with the control VBS01 strain that is not overexpressing SHMT module in the SM medium with and without glycine supplementation, under high CO_2 conditions. **C.** Growth curves of three different serine biosensor strains overexpressing SHMTs along with the control serine biosensor strain not overexpressing SHMT in the SM medium with 250 mM formate under 10% high CO_2 atmosphere in the presence and absence of glycine.
SM - Synthetic minimal medium, SM always contains 1xYNB and 100 mM glucose unless stated otherwise; **YNB** - Yeast Nitrogen Base; **DT** - Doubling Time; **NG** - No Growth; Growth is always measured by absorbance at 600 nm. **SHMTs:** Serine hydroxymethyltransferases. **Green** - different SHMTs; **Ox** - overexpression;

All three S288c_ΔS strains with *SHMTox* module grew well in SM medium supplemented with glycine and formate (**Figure 6C & 6A**). This indicates that glycine conversion to serine did enable growth in presence of formate thus all the auxotrophies are satisfied in these strains. The VBS01 strain did not grow with glycine addition in presence of formate, which is indicating that the control strain (VBS01) without *SHMTox* might not be able to convert glycine to serine to satisfy all the auxotrophies hence growth maybe not supported when SHMTs are not overexpressed. As expected, when glycine is not provided in the medium along with formate no growth is observed from any of the ΔS strains (**Figure 6C**) (Note: glucose is present as the primary carbon source in the all the media conditions explained in the experiment).

VBS06 grows faster, when glycine and formate were provided in the medium, with a doubling time of 3.8 h compared with VBS04 and VBS05 that are growing at 7.9 h and 6 h of doubling time, respectively. In both scenarios, VBS06 with *EcSHMT* showed faster growth compared to *SHM1* and *SHM2* expression VBS04 and VBS05 strains (**Figure 6B & 6C**). Herewith the finding described here, I successfully constructed three ΔS strains of S288c able to convert glycine to serine in the presence of formate – an important pre-requisite for implementation of the RGP. Further steps were taken to engineer the glycine synthesis module to produce glycine from formate and CO_2 to achieve the serine synthesis via the RGP.

5.1.4. Expression of glycine synthesis module of the RGP

The glycine synthesis module can functionally be divided into two submodules: 1. The C1 maturation module – in which formate is ligated to THF and further reduced to CH_2-THF by trifunctional Mis1p in mitochondria. 2. The glycine synthase complex (GSC), which is also confined to mitochondria and metabolically connected with mitochondrial C1 maturation carried out by Mis1p. The GSC complex is composed of Gcv1p, Gcv2p, Gcv3p, and Lpd1p required for glycine biosynthesis from formate in the presence of a high CO_2 atmosphere. The mitochondrial GSC assimilates the methyl group from mitochondrial CH_2-THF and NH_3 from ammonium sulfate and CO_2 to produce a glycine molecule. The MIS1, GCV1, GCV2, and GCV3 were overexpressed under the control of constitutive promoters of *RPL3*, *PYK1*, *TEF1*, and *TEF2*, respectively, to establish the glycine synthesis module. Lpd1p is also involved in the GSC but is not overexpressed to prevent unwanted imbalances that may cause growth inhibition as Lpd1p is involved in other intracellular enzyme complexes.

As described in section 4.5 (**Table 2**), different serine biosensor strains were constructed by expressing glycine biosynthesis modules either on a plasmid (pFM340/pFM345) or from the genome (ChrX-3). As the preliminary experiment, I tested several ΔS strains, i.e., VBS01, VBS02, VBS04, VBS05, VBS06, VBS07, and VBS12 strains (**Table 2**) with/without

overexpression of glycine synthesis or the entire serine synthesis module, for glycine and serine synthesis from formate. All the strains failed to show any growth in SM medium supplied with glucose and formate even with prolonged incubation under high CO_2 conditions. Without overexpressing the complete serine synthesis module (glycine synthesis module + glycine to serine conversion module), no growth is observed.

Efficient glycine to serine conversion is achieved as described in section 5.1.3. In order to achieve serine synthesis from formate, the VBS04, VBS05, and VBS06 strains were transformed with the pFM340 plasmid carrying the glycine synthesis module. This further creates three different ΔS strains, i.e., VBS08 (S288c_ΔS-SHM1-pFM340), VBS09 (S288c_ΔS-SHM2-pFM340), and VBS10 (S288c_ΔS-EcGlyA-pFM340) expressing entire serine synthesis module which includes glycine synthesis module and glycine to serine conversion module. (**Note:** the glycine synthesis module is expressed from the pFM340 plasmid, and SHMTs (glycine to serine conversion module) are overexpressed from the genome) (**Table 2**). Even though Ade3p is involved in cytoplasmic serine synthesis, it was not overexpressed as it is natively active. Then, Adaptive Laboratory Evolution (ALE) was conducted using VBS08, VBS09, and VBS10 strains to achieve glycine and serine synthesis from formate.

5.1.5. ALE for glycine and serine synthesis

ALE was conducted in synthetic minimal (SM) medium and synthetic complete (SC) medium. SC is a semi-rich medium that contains a mix of all amino acids except serine and glycine, whereas SM is a minimal medium that does not contain amino acids. 1x YNB, a blend of essential salts, minerals, and vitamins, is used in both SC and SM media. 100 mM glucose, 250 mM formate, and 100 mM $(NH_4)_2SO_4$ are used in both media conditions. A high formate concentration, 250 mM, was used to drive the biosynthesis of CH_2-THF. A high $(NH_4)_2SO_4$ concentration (100 mM) is provided to drive the amino-methyl transferase reaction on Gcv3p, which is the first reaction in the GSC that is carried out by Gcv1p. Subsequently, a CO_2 molecule is attached to the amino-methyl group releasing a glycine molecule and leaving Gcv3p in the oxidized state. This reaction is carried out by Gcv2p, which reaction requires high CO_2 conditions to operate in the glycine synthesis direction, assimilating a CO_2 molecule. Thus, cultures are essentially incubated in a high CO_2 atmosphere. In the next step of GSC, Gcv3p is recycled for the next round of glycine synthesis by reducing Gcv3p. This reaction is carried out by native Lpd1 activity by investing an NADH. All the concentration values of the substrates, formate, CO_2, and $(NH_4)_2SO_4$, were provided in high concentrations to facilitate better kinetics to drive the RGP but not detrimental to the yeast strains. Hence, I proceeded

with ALE in the explained conditions using serine biosensor strains expressing the entire serine synthesis module, i.e., VBS08, VBS09, and VBS09 strains (**Table 2**).

Table 3: List of strains and media conditions used for ALE
SC and SM media conditions used for testing glycine and serine synthesis in VBS08, VBS09, and VBS10 strains; DO-Ser-Gly is a mix of all the amino acids in recommended quantities for optimal growth of yeast, except serine and glycine. DO- dropout; SC- synthetic complete; SM- synthetic minimal; YNB- Yeast Nitrogen base;

Media →	SC=1x YNB, DO-Ser-Gly, 100 mM $(NH_4)_2SO_4$,100 mM glucose			SM =1x YNB, 100 mM $(NH_4)_2SO_4$, 100 mM glucose		
Strains ↓	Growth positive control	Test glycine and serine synthesis	Growth negative control	Growth positive condition	Test glycine and serine synthesis	Growth negative control
VBS08	SC-Ser+Gly, formate	SC-Ser-Gly, formate	SC-Ser-Gly	SM, Glycine, formate	SM, formate	SM
Growth	+	-	-	++	-	-
VBS09	SC-Ser+Gly, formate	SC-Ser-Gly, formate	SC-Ser-Gly	SM, Glycine, formate	SM, formate	SM
Growth	+	-	-	++	-	-
VBS10	SC-Ser+Gly, formate	SC-Ser-Gly, formate	SC-Ser-Gly	SM, Glycine, formate	SM, formate	SM
Growth	++	+	-	++	-	-

5.1.6. ALE in semi-rich medium

The primary transformants of VBS08, VBS09, and VBS10 strains, carrying both glycine and different serine synthesis modules, were inoculated in SM and SC media under selective conditions. As indicated in **Table 3**, three different media conditions of SM medium and SC-Ser-Gly medium were used. The VBS08 and VBS09 strains did not grow if glycine was not provided in both SC and SM media types.

Surprisingly, the VBS10 strain showed a slight growth to ~0.5 OD_{600} in SC-Ser-Gly medium supplemented with formate but not in minimal media condition unless glycine is supplied (**Figure 7B**). This was the first indication of glycine and serine synthesis from formate in the VBS10 strain and the respective culture was considered as evolution round 01 (Ev01), (**Figure 7A**). The grown cells of VBS10 strain in SC-Ser-Gly medium from the EV01 were diluted in the fresh SC-Ser-Gly medium in Ev02 and incubated. After ~35 days of incubation, growth was observed in this evolution round, Ev02. Cells from this culture were transferred to SM medium with formate with and without 0.25 mM glycine (Ev03) (**Figure 7A & 7B**). Growth was

observed without glycine supply in Ev03, which indicates consistent glycine biosynthesis from formate in SM media.

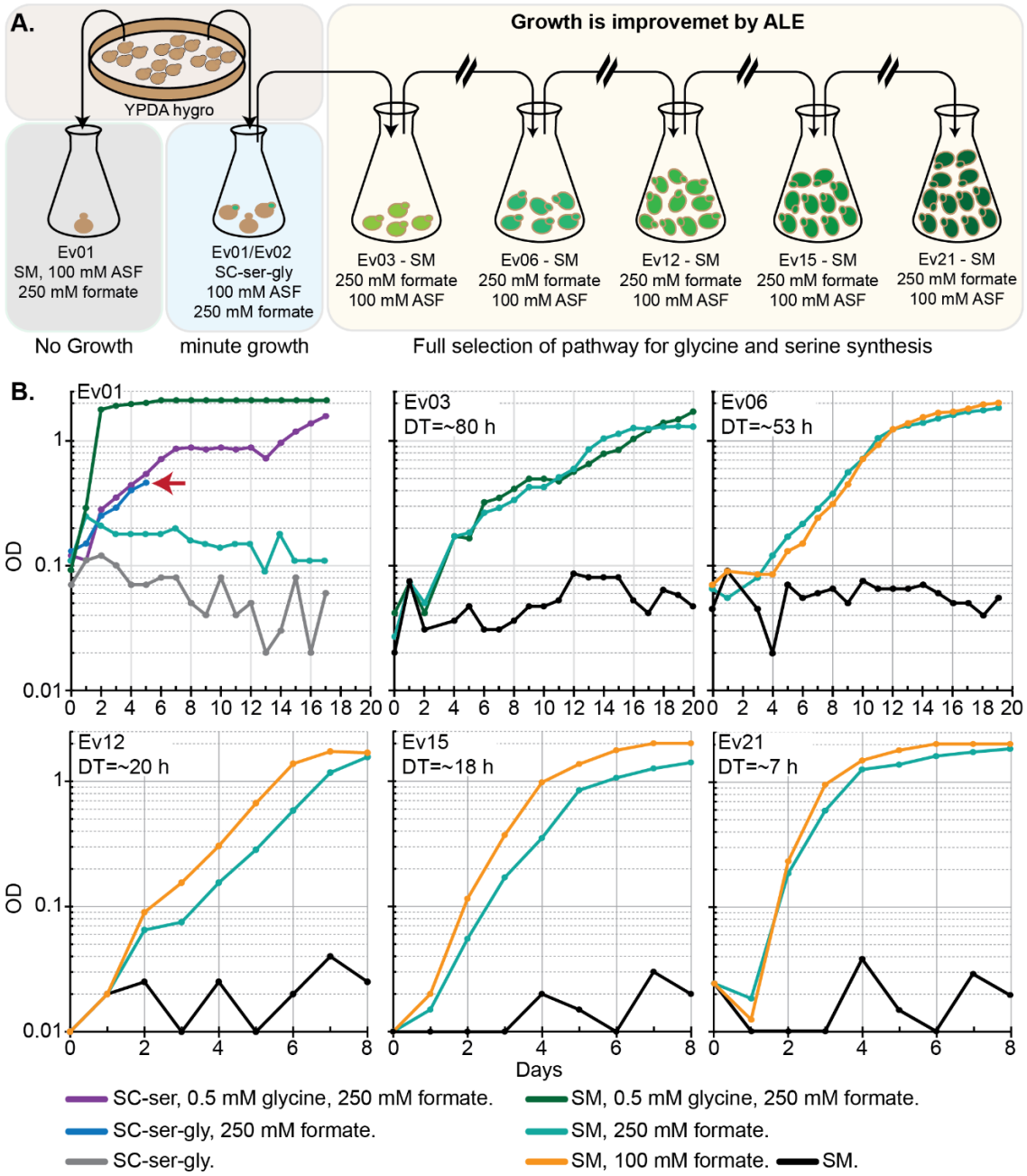


Figure 7: Schematic of ALE process and growth curves of VBS10
A. The schematic of the ALE process of VBS10 in SM and SC-Ser-Gly media conditions supplied with glucose and formate. VBS10 strain is tested for glycine and serine synthesis in SM medium and SC-Ser-Gly medium with 100 mM glucose and 250 mM formate Ev01. Slightly grown culture from Ev01 was sub-cultured further in SC-Ser-Gly medium in Ev02. After 25 days of incubation, grown cells from Ev02 were used as inoculum in SM medium with 250 mM formate and 100 mM glucose in Ev03. Further continued to subculture from exponential phase until Ev21 to improve the growth by improving serine and glycine biosynthesis. **B.** The growth curve of selective intermediate evolution rounds of Ev01, Ev06, Ev12, Ev16, and Ev21 were displayed with a doubling time decreased from 80 h to 7 h in SM medium. **ASF:** Ammonium sulfate ($(NH_4)_2SO_4$); SC-Ser-Gly: Semi-rich medium without serine and glycine.

No growth is observed when formate is not added to the medium (**Figure 7B**). This result indicates that growth of the VBS10 strain is formate-dependent. However, the strain was growing at a doubling time of ~80 h which was very slow growth at this stage of evolution. Further to test if this growth is dependent on high atmospheric CO_2, another round of evolution was conducted in ambient and high CO_2 conditions.

As expected, growth is only observed when cultures were incubated in a high CO_2 atmosphere (**Figure S 2A**), satisfying the requirement of a high CO_2 for the reversal of GSC. At this stage of the evolution experiment, it is evident that growth of the VBS10 strain is high CO_2 and formate dependent, which indicates RGP activity resulting in glycine and serine biosynthesis from formate and CO_2.

ALE was continued as the growth of the strains were still very slow. To improve the growth rate, I continued to sequentially subculture/dilute the cells in fresh medium with formate. In Ev05, growth of the VBS10 strain was tested in two different formate concentrations, 100 mM and 250 mM. Cells reached higher OD when fed with 250 mM formate compared to the lower 100 mM formate condition (**Figure S 2B**). This indicates that a high formate concentration is probably required to drive the RGP to reach higher final ODs. Further evolution was continued in SM medium with 100 mM and 250 mM formate in each evolution round along with a negative control condition, without formate, to monitor if growth was formate dependent. Cells from the 250 mM formate condition were used as seed in every subsequent evolution round to maintain the pressure to drive formate utilization. The 100 mM formate condition was used in parallel, as it is a less toxic condition and to improve the kinetics to operate RGP in lower formate concentration.

In Ev06, a clear growth improvement was observed: The cell population of VBS10_Ev06 were growing at ~53 h of doubling time. At this stage, the 100 mM formate condition was leading to the same final OD_{600} compared to the 250 mM formate condition. During the further evolution, the strain decreased its doubling time to 20 h and 18 h by Ev12 and Ev15, respectively. In round Ev21, the VBS10 cells reached approximately 6-7 h of doubling time in both, 250 mM and 100 mM formate conditions. At this stage, the glucose concentration was decreased to 50 mM, to decrease the main carbon source in order to decrease main carbon source. This 50% decrease in glucose did not affect the growth rates of the strain in Ev21.

For the assessment of the growth improvements of the VBS10 strain during evolution, the Ev03, Ev06, Ev12, Ev16, and Ev21 were selected. Multiple single colonies from the population of each selected Ev were tested in high throughput plate reader experiments. Growth rates of multiple colonies from the population of each selected Evs were assessed and compared statistically (**Figure 8**). Overall significant 8-fold growth improvement was observed from 0.009

h^{-1} to ~0.085 h^{-1} between Ev03-Ev21. Growth improvement between Ev16 and Ev21 is insignificant, but the variance of growth rates among the Ev21 and Ev16 clones is decreased from $1.04x10^{-4}$ to $1.05x10^{-5}$.

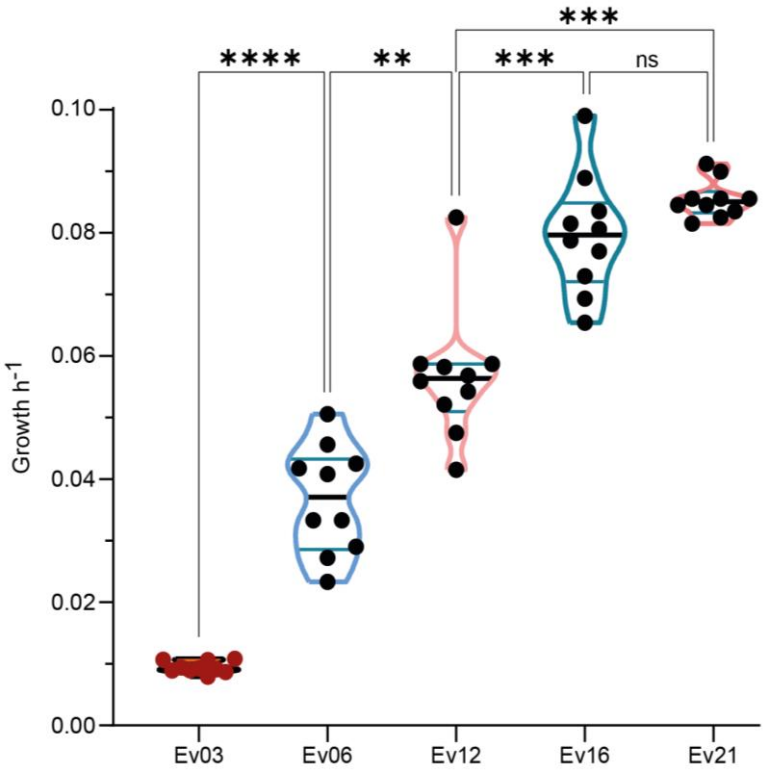


Figure 8: Statistical analysis of growth improvement of VBS10 strain with ALE
Growth rates of multiple colonies of the evolved VBS10 strains assessed. Violin plots are prepared with the distribution of growth rate values of the multiple singles colonies from each evolution selected evolution rounds, i.e., Ev03, Ev06, Ev12, Ev16, and Ev21. One-way ANOVA test is performed to assess the significance of the growth rate improvements and the p-value is calculated. The significance of the growth improvement is indicated with stars based on p -values. If p=<0.0001, it is the highest significance indicated as ****, *** indicates the p=0.0002-0.0005, p=0.0014 is indicated as two **, and p=0.0338 is one *, and no significance was indicated as 'ns.' (Table S3 is can be referred for more details on statistical analysis); Ev- evolution round;

So far, I was able to establish the formate and high CO_2 dependent growth in the VBS10 strain carrying *E. coli* variant of *EcSHMTox* module on genome, and glycine synthesis module on the plasmid (Map S5). Further, it requires the validation of glycine and serine synthesis from formate by carbon tracking experiments.

5.1.7. Validation of serine and glycine synthesis via ^{13}C labeling

Carbon tracing experiments were conducted with the evolved VBS10 strains from Ev10, Ev16, and Ev21 to validate that the glycine and serine synthesis is from formate. These experiments

were conducted in the SM medium using 40 mM glucose as the primary carbon source and 40 mM $(NH_4)_2SO_4$ as the nitrogen source. Single colonies of evolved VBS10 strain from the selected evolution rounds along with WT control were fed with ^{13}C formate/^{12}C formate and incubated in 10% $^{12}CO_2/^{13}CO_2$. The proteinogenic amino acids from the cell biomass of early stationary phase cultures were analyzed using LC-MS. The data was analyzed for the $^{12}C/^{13}C$ ratios in proteinogenic glycine and serine in the evolved VBS10 strain and compared with the WT control. Four different media conditions were adapted to trace formate and CO_2 assimilation via the RGP i.e., ^{13}C formate & $^{12}CO_2$, ^{12}C formate & $^{13}CO_2$, ^{13}C formate & $^{13}CO_2$, and ^{12}C formate & $^{12}CO_2$.

The evolved VBS10 and S288c_WT strains were fed with 250 mM ^{13}C formate and 10% $^{12}CO_2$ in the minimal medium in the first culture condition. Single ^{13}C carbon from formate ligated and reduced to $^{13}CH_2$-THF assimilated via GSC, resulting in single labeled glycine (**Figure 9A; panel 1**). Subsequently, another ^{13}C from formate in the form of $^{13}CH_2$-OH molecule produced from $^{13}CH_2$-THF and water is attached to single labeled glycine leading to twice labeled serine (**Figure 9A; panel 1**). Hence, the labeling data analysis shows that more than 95% of the glycine carries a ^{13}C carbon label, and more than 90% of the serine carries two ^{13}C carbons in the Ev10, Ev16, and Ev21 strains (**Figure 9B; panel 1**). As a control, the biomass of the WT strain, grown in the same culture conditions, did not show similar labeling in glycine and serine. Only a negligible percentage of glycine carry labeled carbon; in incase of serine, 20% of single labeled serine is present. This is because Shm1p/Shm2p assimilates $^{13}CH_2$-THF molecules produced from ^{13}C formate and glycine produced from threonine cleavage to make a pool of serine carrying a single ^{13}C carbon. There is no selection pressure in WT strain as the serine, glycine, and C1 metabolism are not insulated, and also, the serine synthesis module of the RGP was not overexpressed in the case of WT. Hence, the WT strain does not produce glycine from formate.

In the second culture condition, where the evolved VBS10, and S288c_WT strains were fed with 250 mM ^{12}C formate and 10% $^{13}CO_2$, in the minimal medium. The labeling data show that 40% - 50% of glycine and 40% – 48% of serine carry a single ^{13}C carbon, and the rest of the rest carry ^{12}C carbon. Due to the competitive inhibition between the atmospheric CO_2 and glucose-derived intracellular CO_2, only one glycine molecule out of two carries an atmospheric CO_2 molecule (**Figure 9A & 9B; panel 2**). This glycine is converted to serine by assimilating another ^{12}C formate molecule; thus, 50% serine carries only one ^{13}C carbon. The WT strain shows less than 20% of glycine and serine single labeled in the same condition. Due to anaplerotic CO_2 assimilations, 10-20 % of glycine and serine carry single ^{13}C carbons.

In the third culture condition, where the strains are fed with ^{13}C formate and $^{13}CO_2$, 95% of glycine carries a single carbon label, and 40% - 60% carries two labeled carbons due to atmospheric $^{13}CO_2$ assimilation via the RGP as indicated. More than 90% of serine is carrying two labeled carbons from formate as indicated, and 35% - 60% of serine is carrying three labeled carbons due to assimilation of $^{13}CO_2$ from the atmosphere (**Figure 9A & 9B; panel 3**).

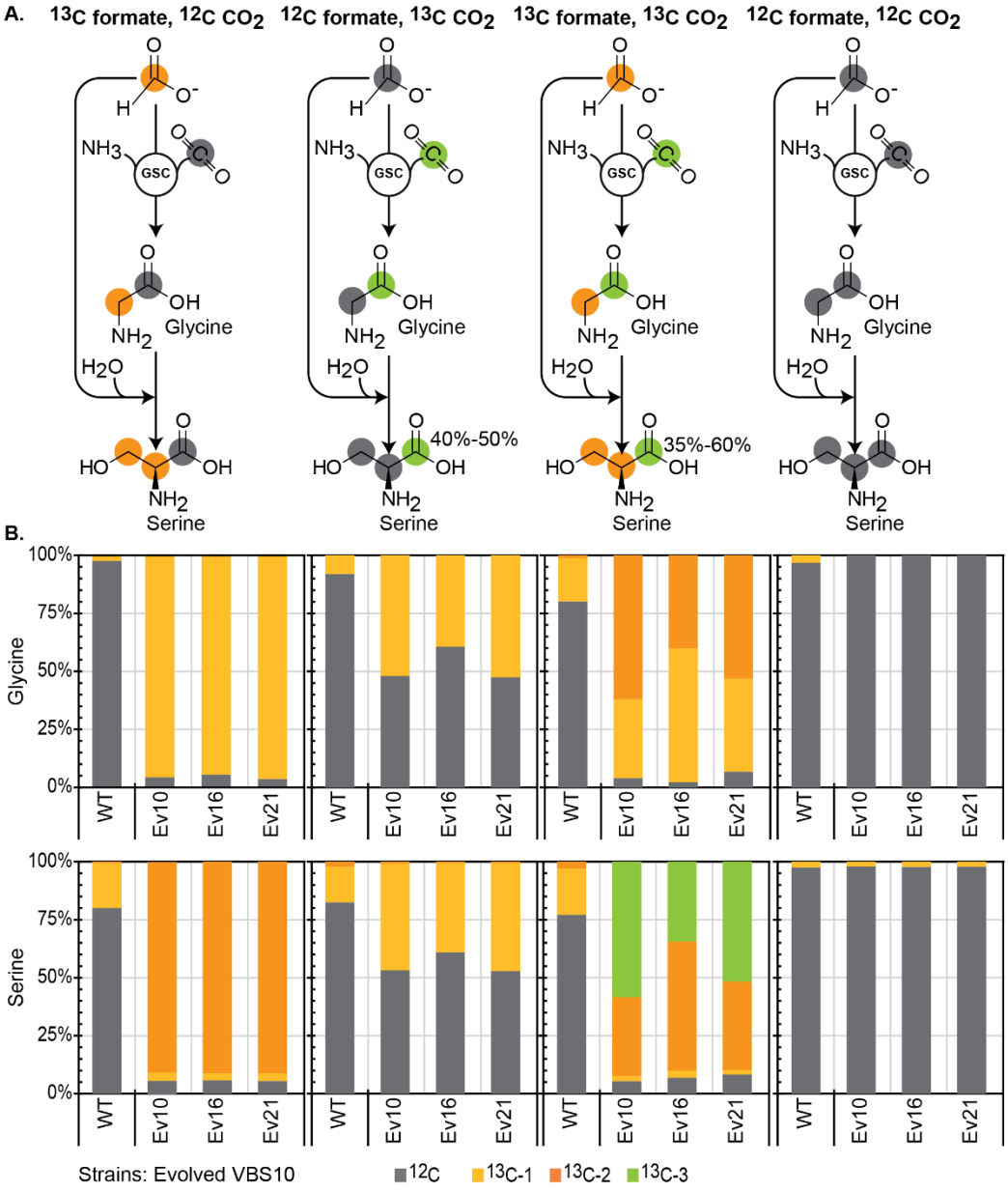


Figure 9: **Validation of glycine and serine synthesis in the evolved VBS10 strains via ^{13}C tracing**
^{13}C labeling data of proteinogenic amino acids glycine and serine from biomass of evolved VBS10 strains from Ev10, Ev16, and Ev21 against WT fed with ^{13}C formate and $^{13}CO_2$. **A.** Panel 1-4 are the metabolic schemes of ^{13}C carbon distribution into glycine and serine produced via the RGP in the given culture conditions. **B.** Panel 1 - 4 represents ratios of ^{12}C and ^{13}C carbons in glycine and serine when the strains are fed with ^{13}C formate & $^{12}CO_2$, ^{12}C formate & $^{13}CO_2$, ^{13}C formate & $^{13}CO_2$, and ^{12}C formate & $^{12}CO_2$ respectively.

In this condition, the WT strain also shows 20% -25% of glycine and serine carrying single labeled carbon, and this can be due to the combined activity of anaplerotic $^{13}CO_2$ assimilation and $^{13}CH_2$-THF assimilation activity of SHMTs.

In the fourth condition, where the strains were fed with only ^{12}C formate and $^{12}CO_2$, no labeled carbons were observed, except noise/contamination of a single carbon label at a negligible level, below 2-3% (**Figure 9A & 9B; panel 4**). Overall, the labeling results confirm that 90%-95% of serine and glycine are derived from the formate in contrast to the WT control. This result validates the establishment of the RGP for glycine and serine biosynthesis from formate and CO_2 via the RGP.

Nonetheless, the experimental results with $^{13}CO_2$ reveal that only approximately 50% of glycine is synthesized via the atmospheric CO_2 assimilation, and the rest of glycine has resulted from the assimilation of intracellular CO_2 derived from glucose. Hence only one molecule of atmospheric CO_2 is incorporated for every two cycles of reverse GCV activity due to the mutual and competitive inhibition of glucose-derived CO_2 and atmospheric CO_2 molecules in the tested conditions.

5.1.8. Evolution in glycine limiting conditions

The evolution of VBS08 and VBS09 strains was also conducted simultaneously with the VBS10 strain in a different approach. The *EcSHMT* overexpression was probably able to create enough intracellular glycine requirement in VBS10, which might have triggered the glycine synthesis from formate, thus resulting in a slight growth in SC-Ser-Gly medium supplemented with formate (**Figure 7B**). This growth is further improved with evolution as explained in the previous section 5.1.6. This was not seen in the case of VBS08 and VBS09 strains that carry *ScSHM1* and *ScSHM2* modules, respectively. I assumed that the overexpression of *ScSHMTs* and their enzyme activity might not have been able to create enough intracellular glycine requirement/limitation, hence not triggering the formate conversion to glycine. To create artificial/synthetic glycine limitation, I followed an approach to expose the cells to glycine limiting conditions.

5.1.9. Process of evolution glycine limiting conditions

As it is explained in section 5.1.6., the primary transformants of VBS08 (S288c_ΔS-SHM1-pFM340), VBS09 (S288c_ΔS-SHM2-pFM340), and VBS10 (S288c_ΔS-EcGlyA-pFM340) strains were inoculated in SM and SC media under selective conditions. As it is indicated in table 3, the VBS08 and VBS09 strains did not grow in SM or SC medium without glycine supply, unlike the VBS10 strain. However, the VBS08 and VBS09 strains were able to grow

well in the SM medium supplemented with formate and glycine as secondary carbon sources. Hence, evolution for these strains was conducted in the SM medium with glucose and formate in synthetic limiting glycine conditions to evolve the strains to operate RGP for glycine and serine synthesis. In this evolution approach, the glycine concentration in the medium was decreased gradually with each subsequent subculture and the evolution process was continued under consistently limited glycine conditions. If glycine and formate are available in the medium, formate drives the synthesis of CH_2-THF, which, together with the glycine provided in the medium, can be driven for serine synthesis. In this scenario, glycine utilization for the synthesis of C1 units would be inhibited/reduced as high formate concentration is likely to provide strong driving force to produce all intracellular C1 units in the serine biosensor strain. Hence, glycine would be assimilated for serine synthesis, which is a parallel flux for RGP flux. As glycine is consumed for biomass and serine synthesis, its level depletes. Keeping the strains repeatedly in this condition while gradually decreasing the glycine concentration with every subsequent passage might trigger the glycine synthesis from formate.

To test this hypothesis, both strains, VBS08 and VBS09, grown on SM medium with 250 mM formate and 1 mM glycine, are used as inoculum in the Ev02 with 0.5 mM glycine. Subsequently, the glycine concentration was further decreased to 0.25 mM, 0.125 mM, and 0.05 mM in Ev03, Ev04, and Ev05, respectively (**Figure 10A**). Both VBS08 and VBS09 strains were able to grow at a glycine concentration as low as 0.05 mM, thereby reaching reduced final ODs (**Figure S 3A & 3B**). Further, both strains were continuously cultured in 0.05 mM glycine and formate, but the VBS09 strain did not grow further in this process (**Figure S 3D**) (reasons were not speculated). Nevertheless, in the Ev06, the VBS08 strain evolved and started to grow without glycine, which is an indication of glycine and serine synthesis from formate (**Figure 10B**).

However, the VBS08 strain showed signs of growth improvement for a few evolutions rounds but then started to grow in the absence of formate. This indicates that the strain has re-acquired its ability to produce glycine and serine from glucose – thereby breaching the selection scheme (**Figure 4D** and **Figure S 3C**). Simultaneously, the labeling experiment with ^{13}C formate also revealed that 35% - 60% of glycine and serine are synthesized from glucose but not from formate, which was a major leakage in the VBS08 strain (**Figure 4D**).

To conclude, the VBS10 strain overexpressing *E. coli* SHMT engineered in the background of the S288c strain was successfully evolved for glycine and serine synthesis via the RGP. In contrast, the VBS08, and VBS09 overexpressing *SHM1* and *SHM2* modules, respectively, constructed in the same strain background strain S288c, failed to evolve.

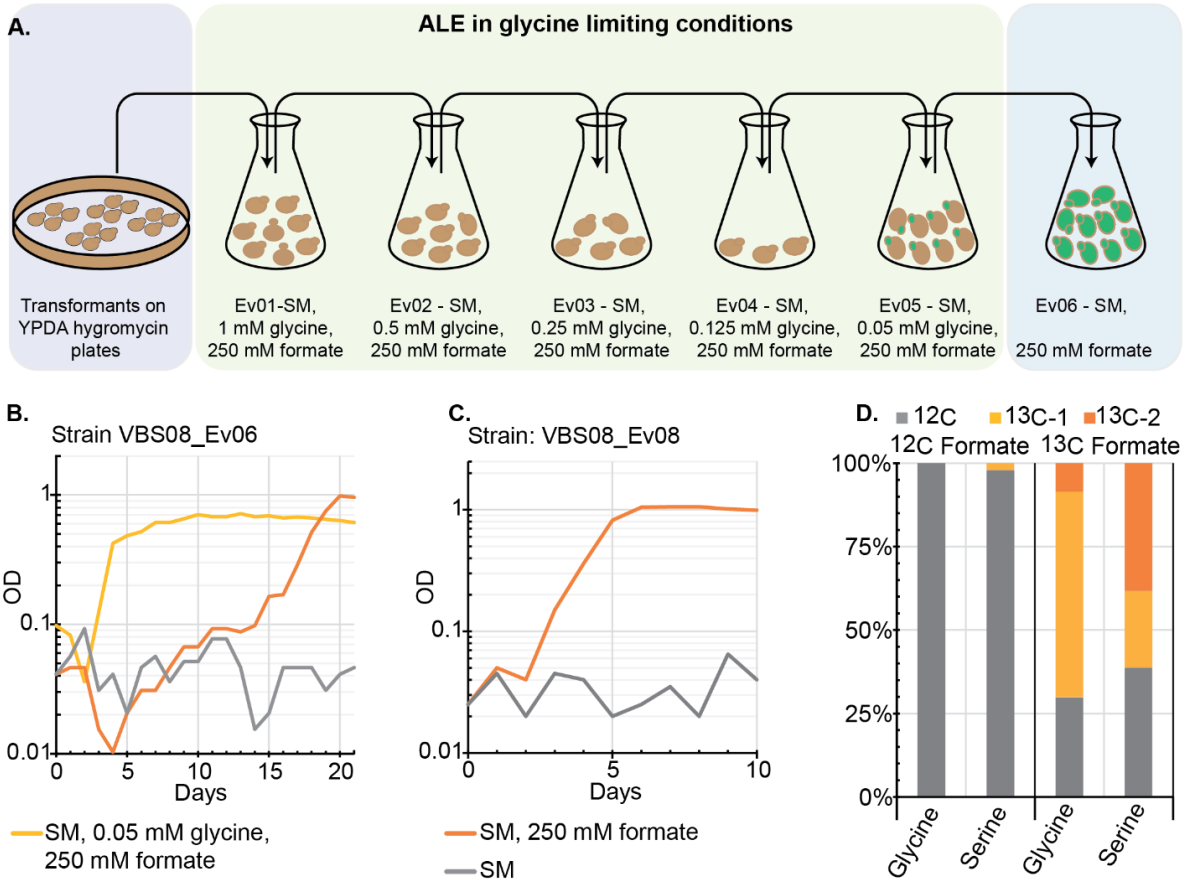


Figure 10: Adaptive laboratory evolution of VBS08 and VBS09 strains

A. Graphical schematic of the adaptive laboratory evolution conducted for VBS08 and VBS09 strains. The scheme indicates the complete glycine limiting process beginning from Ev01 to Ev06. **B.** Growth curves of VBS08 strain in Ev06 in SM medium with formate in the presence or absence of glycine and in SM control. **C.** Growth curves of the VBS08_Ev08 strain in the SM medium with and without formate. **D.** Ratios of ^{12}C and ^{13}C carbons are shown in proteinogenic glycine, and serine. **SM**= 1x YNB, 100 mM glucose, and 100 mM $(NH_4)_2SO_4$.

5.2. Establishing the RGP in *S. cerevisiae* FL100

S. cerevisiae FL100 is another prototrophic strain used in this study to establish the RGP. It is also a laboratory strain.The genome sequence-based phylogenic tree, this yeast is placed close to Fosters (B & O) *S. cerevisiae* strains used in commercial beer production and distant from traditional laboratory strains of S288c[107,108]. Unlike S288c, FL100 carries Lsg-TRR (codes for transcription regulatory protein highly expressed in fermentation conditions) like genes similar to industrially employed Fosters B & O strains used for commercial production[107]. As part of the multiple strain approach to establish the RGP, the FL100 variant was also employed in this study.

5.2.1. Construction of serine biosensor strains in FL100

The construction of the serine biosensor strain in the S288c background was already described in the previous section 5.1.1 of this chapter, and it was also shown that the strains can covert glycine to serine and vice versa. In the case of the FL100 strain, a slightly different approach was employed: Two different serine biosensor strains were constructed, namely VBS13 (FL100 *agx1 gly1 ser1 shm1*) and VBS14 (FL100 *agx1 gly1 ser1 shm*2). The VBS13 strain carries *SHM1* knockout along with the *SER1*, *GLY1*, and *AGX1* knockouts, in which the glycine conversion to serine is restricted to the cytoplasm (**Figure S 4A**). The VBS14 strain carries *SHM2* knockout additionally along with the *SER1*, *GLY1*, and *AGX1* deletions; hence, serine synthesis in this strain is confined to the mitochondrial compartment (**Figure S 4B**).

Phenotype validation experiments with these strains were conducted in SM medium with combinations of glycine, serine, and formate supplementations. Glucose was used as the primary carbon source in the medium. Both VBS13 and VBS14 strains did not grow when only glycine was supplemented in the SM medium, unlike the VBS01 strain. However, when both glycine and formate were added to the medium, growth was enabled in both FL100_ΔS strains. In contrast to the VBS01, these two FL100_ΔS strains were able to convert glycine to serine in the presence of formate without overexpression of the SHMT module (**Figure S 4A & 4B**). Serine supplementation to the medium enabled growth of the VBS13 strain in which serine can be converted glycine and C1 unit in the cytoplasm by Shm2p. In contrast, the VBS14 strain was not able to grow when serine was supplemented in the medium; this could be because of the less overall activity of Shm1p compared to cytosolic isozyme, which was knocked out in VBS14. As expected, the supply of both glycine and serine rescued growth in both strains in the presence or absence of formate (**Figure S 4A & 4B**).

5.2.2. ALE of FL100 for glycine and serine synthesis from formate

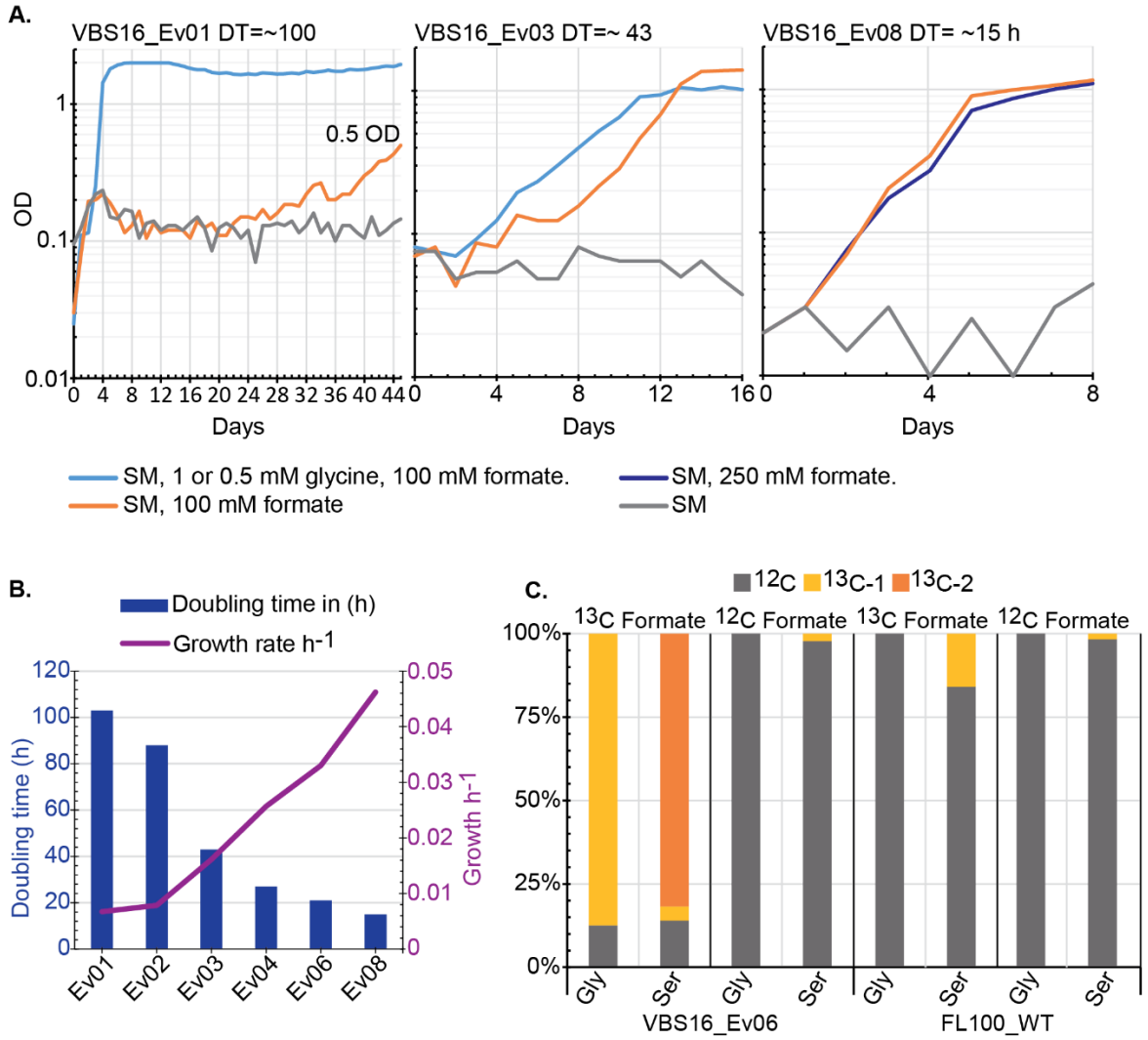


Figure 11: Adaptive laboratory evolution of the VBS16 strain for glycine and serine synthesis
A. Growth curves of VBS16 strain in SM medium supplemented with formate in the presence and absence of glycine and SM medium without formate from the beginning to Ev08. **B.** Doubling time and growth rates of the VBS16 the from Ev01 to Ev08. **C.** Ratios of ^{13}C and ^{12}C carbon in the amino acids from the biomass hydrolysate of the VBS16 strain. **DT** - doubling time. **SM**= 1x YNB, 100 mM glucose, and 100 mM $(NH_4)_2SO_4$.

Glucose was used as the primary carbon source in the minimal medium. As expected, growth was enabled in both strains when the medium was supplemented with 1 mM of glycine. Interestingly, the VBS16 (VBS14+pFM340) strain, in which serine synthesis is confined to mitochondria, grew up to 0.5 OD in 45 days of prolonged incubation under a high CO_2 atmosphere in the SM medium with 100 mM formate without glycine (**Figure 11A**). In contrast, the VBS15 (VBS14+pFM340) strain, in which serine synthesis is confined to cytoplasm, is unable to grow even with prolonged incubation.

Thus, the VBS16 strain was sub-cultured in SM medium with formate, and growth was monitored during eight rounds of evolution (**Figure 11A**). A clear growth improvement is observed from Ev01 to Ev08, 0.0067 h^{-1} to 0.046 h^{-1} which is a 5.86-fold growth improvement. In terms of doubling time is decreased to ~15 h from ~100 h (**Figure 11B**).

5.2.3. Validation of glycine and serine biosynthesis from formate

Next, a ^{13}C formate labeling experiment was conducted to validate glycine and serine synthesis from formate. In the biomass of the VSB16_Ev06, more than 85% of glycine carries a single ^{13}C carbon from the formate, and more than 80% of serine carries two ^{13}C carbons. (**Figure 11C**). In the FL100_WT, labeling was not observed in glycine, but approximately 20% of serine is single labeled which is due to the reversal activity of the cytoplasmic Shm2p activity. As expected, no labeling pattern was observed when ^{12}C formate was fed to the strain. Thus, the labeling data validates that a substantially large amount of glycine and serine are produced from formate even without overexpression of the SHMT module in the VBS16 strain overexpressing only the glycine synthesis module.

Nevertheless, at this stage, between both evolved strains, the VBS10 strain of S288c and VBS16 of FL100, the former was the most promising and advanced with respect to serine synthesis and growth rates. The doubling time of the VBS10-Ev21 strain was 6-7 h, whereas VBS16_Ev08 is 15 h. VBS10_Ev21 strain also reached higher final ODs compared to VBS16_Ev08. Hence, I carried the VBS10_Ev21 strain forward to characterize and evolve further. While continuing the further evolution of VBS10_Ev21 in glucose-limiting conditions, I characterized the effects of substrates to produce RGP-dependent biomass, and whole-genome sequencing was conducted to identify the mutations that occurred during the evolution to improve the serine synthesis from formate.

5.3. RGP-dependent biomass production of VBS10_Ev21

As described in the previous sections, I successfully constructed serine biosensor strains, established the serine and glycine synthesis via RGP, evolved to improve the pathway activity during the evolution. I then validated the RGP activity for glycine and serine synthesis from formate and CO_2 through carbon tracing experiments.

During RGP-dependent growth, the serine, glycine, C1 units, and associated pathway requirement are satisfied from formate, while the remaining biomass is covered by glucose. For the evolution process, all three substrates of the RGP were provided in the elevated concentrations, i.e., 10% CO_2 atmosphere, 250 mM formate, and 100 mM $(NH_4)_2SO_4$, to provide driving forces at non-lethal substrate concentrations to enable the RGP activity. In the course of evolution, enhanced growth rates indicate the improvement of the RGP activity in the evolved VBS10 strains. Thus, to understand the dependency of each substrate provided for the production of RGP-dependent biomass, I characterized the VBS10_Ev21 strains using high-throughput plate reader experiments providing various concentration of each substrate.

5.3.1. Formate gradient and the RGP dependent biomass production

A wide range of formate concentrations between 2-1000 mM was tested in VBS10_Ev21 strains using growth as a readout to check the effect of formate concentration (**Figure 12A-B**). Growth of the VBS10_Ev21 strains was observed with supplementation of formate as low as 10 mM. This is an indication of the high sensitivity of the RGP pathway in the evolved VBS10 strains though the growth is limited at this low concentration. Nevertheless, supplementation of 25 mM formate in the medium improved growth substantially, even though growth rates and final OD_{600} are slightly reduced as compared to the 50 mM and 75 mM formate concentrations. Best doubling times are observed from VBS10_Ev21 strains in the formate concentrations between 50 mM to 125 mM, which is 7 h. At the same time, 200 mM and 300 mM formate concentrations resulted in the slightly lower final biomass with a slight concentration increase in the doubling times (**Figure 12A-B**). Growth is also observed even in 500 mM, 750 mM, and 1000 mM formate concentrations (Figure 12A-B and S5A & S5B); however, severe growth inhibition is observed in high formate concentrations, but still, a prominent pathway activity is observed even in 750 mM formate (**Figure 12B**). This suggests that yeast is highly tolerant of formate, making it a promising host.

Statistical analysis by comparing growth rates of the VBS10_Ev21 in the gradient of formate concentrations shows that the growth of the strain is significantly impaired when supplied with formate below 50 mM and above 300 mM formate (**Figure 12B**).

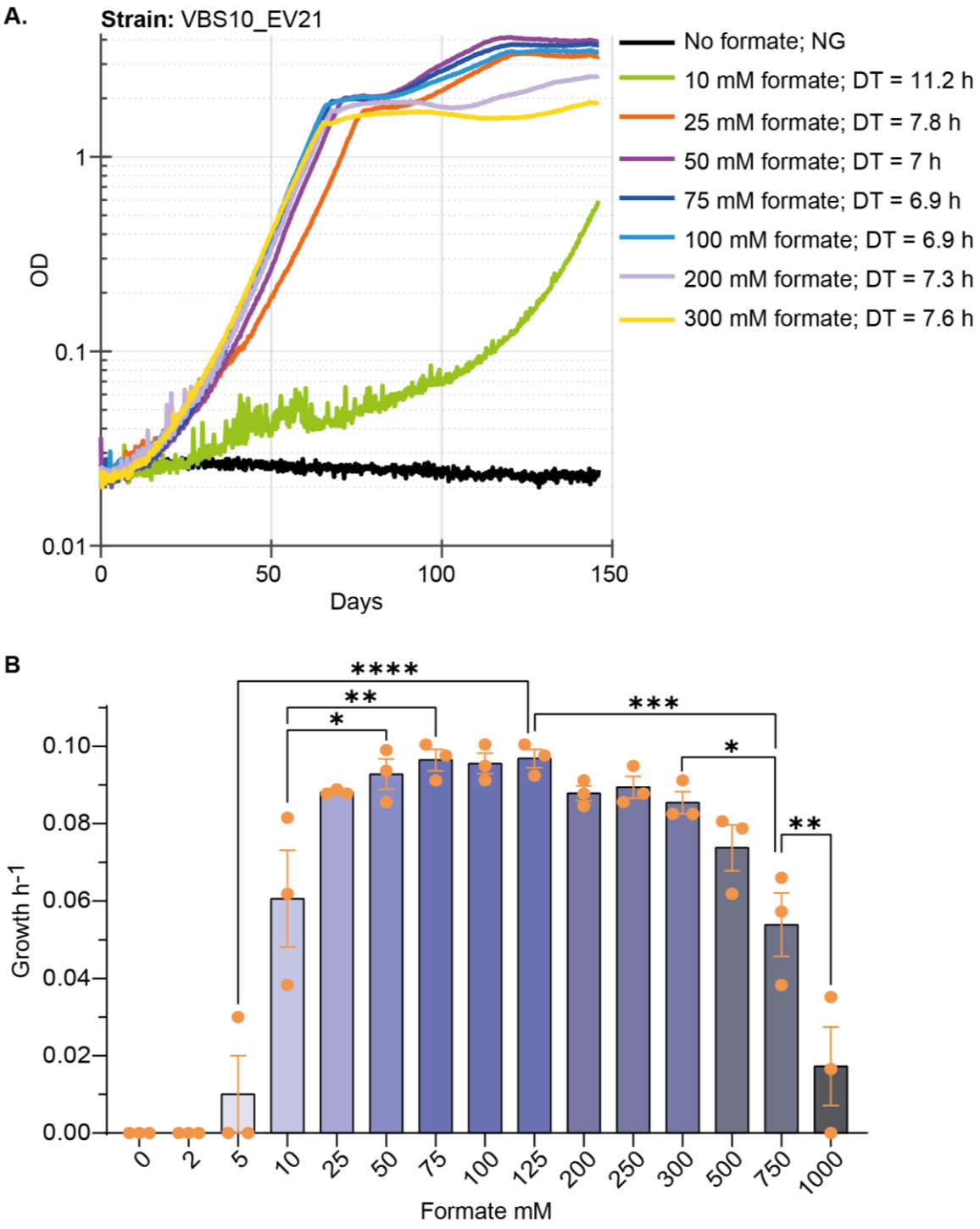


Figure 12: Growth of VBS10_Ev21 strain in formate gradient

A. Growth curves of VBS10_Ev21 with formate concentrations from 10 mM to 300 mM in SM medium with glucose. **B.** Growth rates of VBS10_Ev21 across several concentrations of formate. A one-way ANOVA test and multiple comparisons were performed. Growth limiting or inhibiting concentrations of formate were indicated compared to optimum formate concentration. (Refer to **Table S4** for more details on statistical analysis and p values).

5.3.2. $(NH_4)_2SO_4$ for RGP-dependent biomass production

The methyl group from the CH_2-THF and NH_3 from $(NH_4)_2SO_4$ are together attached to the Gcv3p in the aminomethylation reaction carried out by Gcv1p. To check if varying concentrations of $(NH_4)_2SO_4$ influence the kinetics of this reaction and alter growth of the evolved VBS10_Ev21 strain, I tested a gradient of $(NH_4)_2SO_4$ between 10 mM - 120 mM with a constant 100 mM formate and glucose concentrations in the medium (**Figure 13A-B**).

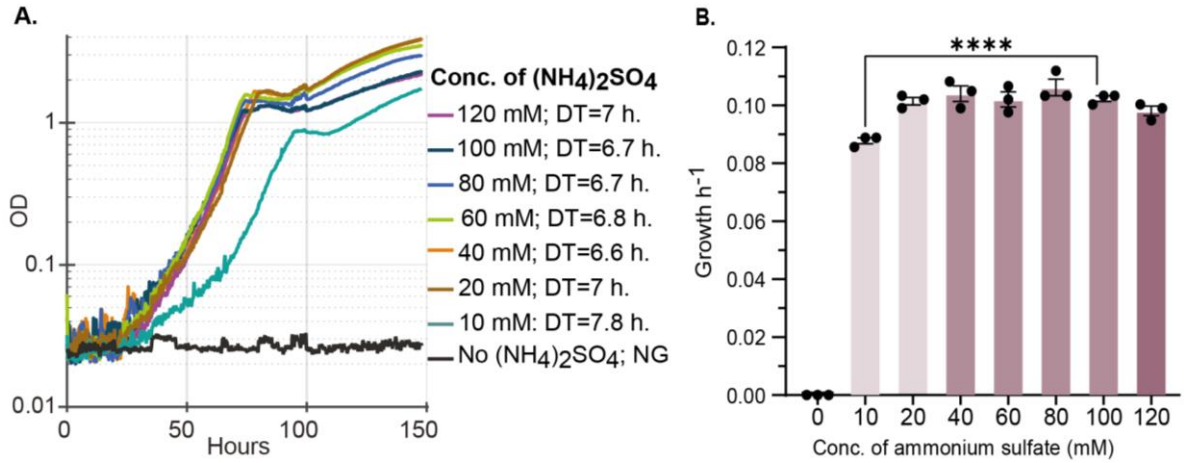


Figure 13: Growth of VBS10_Ev21 strain in ammonium sulfate gradient
A. Growth curves of VBS10_Ev21 with $(NH_4)_2SO_4$ concentrations from 10 mM to 120 mM in SM medium. **B**. Statistical analysis of the growth rates in different tested concentrations of $(NH_4)_2SO_4$ (refer to **Table. S5** for statistical analysis).**SM**= 1x YNB, 50 mM glucose, 100 mM formate.

VBS10_Ev21 and S288c_WT strains were grown in minimal medium supplied with different concentration of $(NH_4)_2SO_4$, i.e., 10 mM, 20 mM, 40 mM (1x), 60 mM, 80 mM, 100 mM and 120 mM. Changing the $(NH_4)_2SO_4$ concentration as a nitrogen source did not influence the growth rates of the VBS10_Ev21 strain indicating that ammonia concentration is not rate-limiting for RGP-dependent biomass production (**Figure 13A**). A slight reduction in growth was observed when 10 mM and 120 mM $(NH_4)_2SO_4$ were provided in the medium (**Figure 13B**). This effect was observed in both S288c_WT and VBS10_Ev21 (**Figure S 6**). All other ammonia concentrations, from 20 mM to 100 mM, did not affect the growth rates of the VBS10 strain or S288c_WT. Thus, ammonia concentration might not have significant influence on the RGP-dependent biomass production in the evolved VBS10 strain (**Figure 13B**).

5.3.3. Influence of CO_2 for RGP-dependent biomass production

Operation of the RGP requires an elevated CO_2 supply to reverse the Gcv2p activity to release glycine from Gcv3p by assimilating a CO_2 molecule. This reaction is the opposite of the natural glycine decarboxylation in ambient atmosphere, leading to the release of CO_2. Hence, to reverse this flux of glycine decarboxylase to glycine synthesis, it is essential to provide a high CO_2 atmosphere. A 10% CO_2 atmosphere is used in the evolution process to achieve serine synthesis via the RGP. To analyze the effect of CO_2 concentration on the glycine synthase activity of the RGP, I tested growth of the VBS10_Ev21 strains under conditions with various CO_2 levels, i.e., 10%, 5%, 2.5%, 1%, and ambient CO_2 condition (**Figure 14**, **Figure S 7**).

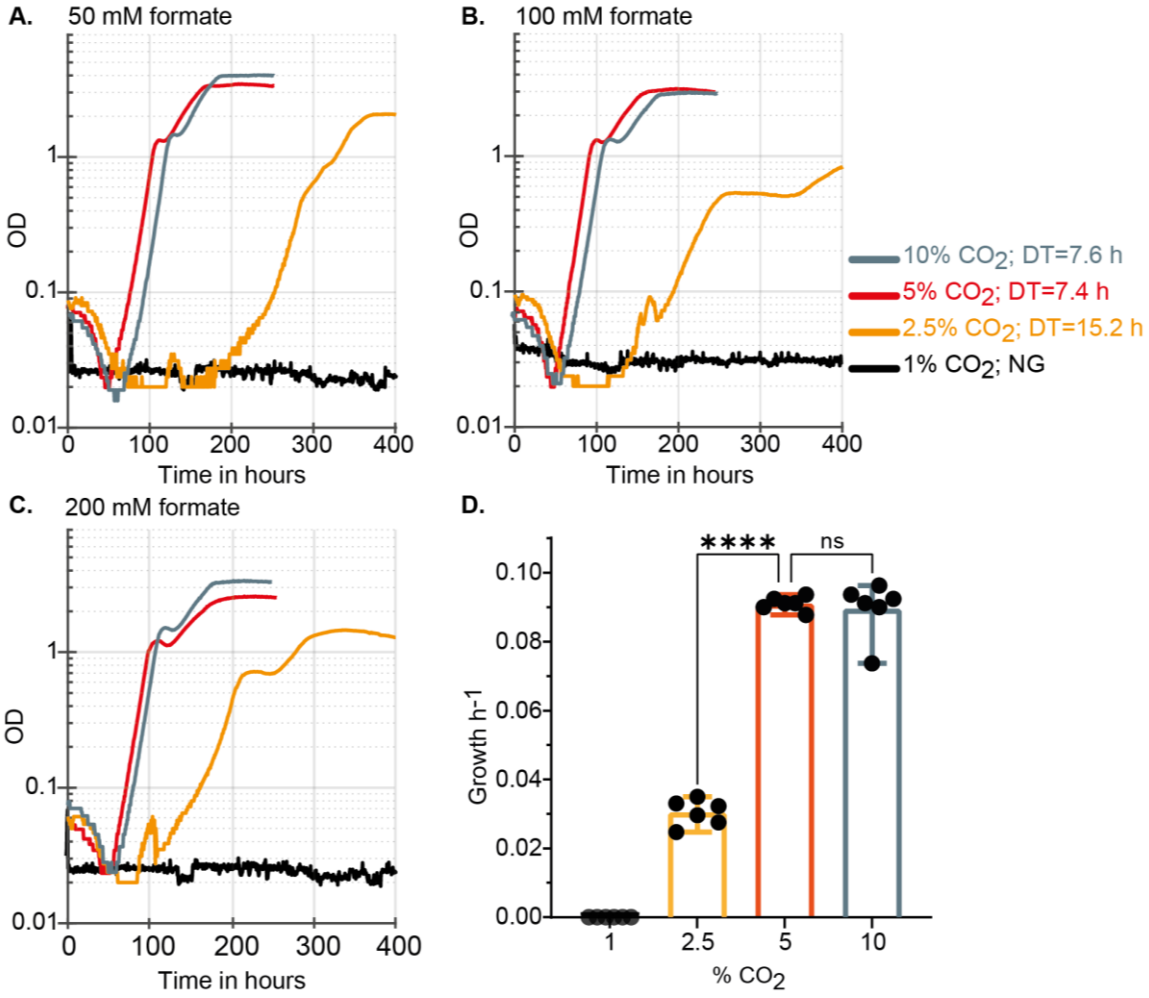


Figure 14: Growth of VBS10_Ev21 strain in the different percentages of atmospheric CO_2
Growth curves of VBS10_Ev21 strains incubated in 1%, 2.5%, 5%, and 10% CO_2 atmosphere supplied with **A.** 50 mM, **B.** 100 mM, **C.** 200 mM of formate concentrations in SM medium. **D.** One-Way ANOVA analysis of CO_2 effect on the growth of VBS10_Ev21 strain. **SM**= 1x YNB, 100 mM $(NH_4)_2SO_4$, 50 mM glucose. (Refer to **Table. S6** for statistical analysis).

As expected, no growth of VBS10_Ev21 is observed in the ambient or 1% CO_2 atmospheres, due to the inoperability of the RGP. In contrast, S288c_WT grows in these two conditions as its growth is not dependent on RGP, and there is no selection; thus, entire biomass can be produced from glucose (**Figure S 7A & 7B**). When the CO_2 level in the incubation atmosphere is increased to 2.5%, growth of VBS10_Ev21 is observed, indicating glycine synthesis through the RGP. However, growth rates in these conditions are poor with increased lag phase time (**Figure 14A-C**). Yet, a decreased lag phase is observed with the increase of formate concentration in this condition. Growth is improved substantially when the strain was incubated under a 5% CO_2 atmosphere. Nevertheless, growth rates are not improved further when the CO_2 percentage is increased from 5% to 10% (**Figure 14A-C**). Statistical analysis of growth from different high-throughput growth experiments in different levels of CO_2 and formate conditions shows that: increasing CO_2 percentage from 2.5% to 5% improved growth significantly from ~0.03 h^{-1} to ~0.09 h^{-1} with a p-value of 0.007 (**Figure 14D**). Hence, a high CO_2 atmosphere plays a significant role while producing RGP-dependent biomass. At least a 5% CO_2 atmosphere is necessary for the optimal growth rate of the VBS10_Ev21 strain.

Even though GSC is functional at 2.5% CO_2, at this stage, it requires further improvement to enhance glycine synthesis at this high CO_2 conditions by improving CO_2 assimilation by Gcv2p. Enzyme engineering or directed evolution can be conducted to improve the CO_2 assimilating ability of Gcv2P. Additionally, synthetic organelle or CO_2 enrichment strategies in mitochondria can be explored to operate the GSC system better in the yeast mitochondria to facilitate CO_2 assimilation at lower or ambient CO_2 atmospheres.

5.4. Genome sequencing of evolved VBS10 strains

The successfully evolved VBS10 strains synthesized glycine and serine from formate and improved their growth in the evolution process from Ev02 to Ev21 while growing on formate and glucose selective media. The VBS10 strains from different stages of evolution were sent for NGS to identify if there are any mutations acquired on the genome in the evolution process. Four best growing clones/clones (biological replicates) tested for growth improvement (**Figure 8**) from each selected evolution rounds, i.e., Ev02, Ev06, Ev16, and Ev21, were sequenced to identify genomic mutations. The genomic DNA of evolved VBS10 strains were isolated and sent for next-generation sequencing (NGS) along with the S288c_WT and the unevolved VBS10 strain as the controls.

The genome sequencing data was analyzed using the UNIX-based Breseq program adapted for the haploid yeast genome (refer to section 4.8). The genome sequence data of S288c_WT, unevolved and evolved VSB10 strains were aligned to reference genome sequence of the

parental S288c. The mutations on the genomes of all the strains predicted from the reference genome sequences were identified and further compared among the WT and unevolved VBS10 strains. All the genome modifications, predicted against S288c reference genome sequence, present on the WT and the unevolved VBS10 were considered as mutations that do not influence RGP-dependent growth. The rest of the predicted mutations that were present only on evolved VBS10 strains were considered as the mutations that may directly or indirectly contribute to the enhanced formate utilization and RGP-dependent biomass synthesis.

A mutation that is present nearly in all reads was considered a true mutation. The mutations present on the genome of at least three clones (as biological replicates) out of 4 from each selected Evs were considered the mutation that may support the RGP-dependent biomass production. The mutations identified were divided into two categories based on the location, i.e., coding region mutations and intergenic mutations. The deletions and the point mutations with an amino acid modification in the CDSs were regarded as the coding-region mutations. The rest of the mutations that are present in the non-coding regions of the genome or on the regulatory elements were considered intergenic mutations.

5.4.1. Coding region mutations

The evolved VSB10 strain acquired mutations in the coding regions of several genes i.e., *GDH1*, *PET9*, *IDH1*, *UTP10*, *SUI2*, and *ASH1* (**Table 4**). (Please refer section 5.1.6 and figure 8 for growth improvement) In the CDS of *GDH1*, a deletion of 108 bp is identified between 223-330 base pairs which are equivalent to 36 amino acids. This mutation was identified in the cells right from Ev01 of the VBS10 strain.

Table 4: List of mutation identified in the coding regions of evolved VSB10 from Ev01-Ev21
Δ- deletion, E.g.: A→G – Example of a point mutation where A is converted to G

Gene	Mutation	Ev02	Ev06	Ev16	Ev21	Annotation	Description
GDH1	Δ108 bp	+	+	+	+	coding (223-330/1365 nt)	glutamate dehydrogenase (NADP(+)) GDH1
PET9	A→G	-	+	+	+	I275T (ATT→ACT)	ADP/ATP carrier protein PET9
IDH1	G→T	-	+	+	+	P299Q (CCA→CAA)	isocitrate dehydrogenase (NAD(+)) IDH1
UTP10	C→A	-	+	+	+	M1I (ATG→ATT) †	snoRNA-binding rRNA-processing protein UTP10
SUI2	G→T	-	-	-	+	R88L (CGT→CTT)	translation initiation factor eIF2 subunit alpha
ASH1	Δ3 bp	-	-	-	+	coding (1237-1239/1767 nt)	DNA-binding transcription repressor ASH1

The *GDH1* gene encodes for Glutamate dehydrogenase 1 (Gdh1p); it is a cytoplasmic enzyme involved in NADPH-dependent synthesis of glutamate from NH_3 and α-ketoglutarate (α-KG)[109,110]. The mutation generates the peptide with the deletion of 36 amino acids between

74-110 amino acids. This domain is involved in cofactor binding and polymerization, loss of which would be expected to result in loss of activity (http://pfam.xfam.org/family/ELFV dehydrog#tabview= tab1).

PET9 is a major ATP/ADP carrier protein present in the mitochondrial inner membrane involved in the translocation of ATP/ADP between cytoplasm and mitochondria[111,112]. Here, a point mutation was identified that resulted in an amino acid modification (Ile275Thr). This mutation is observed in the Ev06 onwards and was retained even later in Ev16 and Ev21. When protein is localized on the mitochondrial inner membrane, this mutation is present in the matrix of the mitochondria. This mutation is not reported earlier, but several amino acid mutations in this region lead to loss of Pet9p function[113].

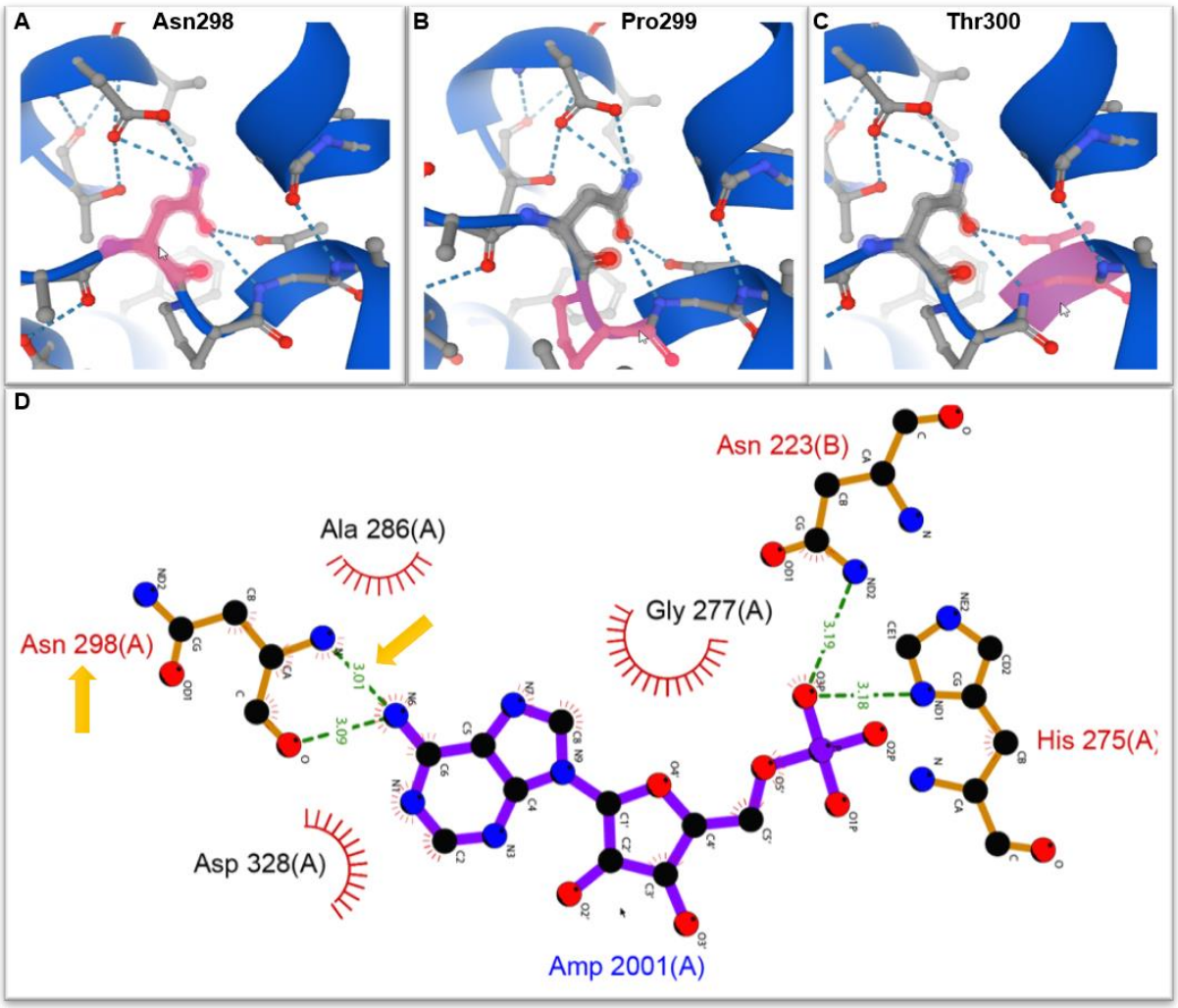


Figure 15: IDH1-P299Q mutation in AMP interacting domain
As highlighted in pink color, the WT Idh1p comprises **A.** Asn298, **B.** Pro299, **C.** Thr300. The mutated Idh1p carries Pro299Gln. **D.** Asn298 interacts with AMP and undergoes conformational changes to activate Idh2p.

A point mutation on *IDH1* was identified from Ev06 and continued to appear in Ev16 and Ev21 also. Iso-citrate dehydrogenase (Idh1p) is an activation part of the Idh1p and Idh2p complex, which carries a catalytic activity site. Based on the intracellular availability of AMP, Idh1p binds to the AMP and undergoes conformational changes to activate the Idh2p, which oxidizes iso-citrate to α-KG releasing NADH[114,115]. In silico analysis using PyMol and a Alphfold online tool suggest that the Asn298-Pro299-Thr300 are involved in the binding to AMP. The mutation Pro299Gln in the AMP binding site changes this domain to Asn298-Gln299-Thr300 (**Figure 15**). It can be speculated that this mutation might be helping to activate the enzyme without binding to AMP, leading to NADH production in mitochondria which may supports the GSC of the RGP. Utp10p is part of the small subunit complex that is involved in the 18s rRNA processing; this mutation appeared from Ev06 and was retained in the later Ev rounds. *SUI2* is the alpha subunit of the translation initiation factor eIF2 that is involved in start codon identification. A point mutation on G→T results in the amino acid change at R88L. A codon deletion was identified in *ASH1*, which codes for a transcriptional repressor. These two mutations occurred in the Ev21 clones but were not found in earlier evolution rounds (referred to https://www.yeastgenome.org).

5.4.2. Intergenic mutations identified

Intergenic mutations appeared between *GSH1/LSB6*, *MET6/IES5*, *VAR1/21s_rRNA*, *ERG26/EFM5*, and *COX3/tM(CAU)Q2* (**Table 5**). Nucleotide changes from A to G appeared in the promoter region of glutamyl cysteine ligase1 (Gsh1p), involved in the catalysis of the first reaction in glutathione biosynthesis. A point mutation changing C to A between *MET6* and *IES5* genes appeared. *IES5* is involved in telomere maintenance. Met6p is involved in the methylation of homocysteine in the final reaction of methionine biosynthesis, in which a CH_3-THF is assimilated (referred from https://www.yeast-genome.org/). These two mutations appeared right from the Ev02 and carried through the evolution. These mutations are of interest because their early appearance in evolution and the functional connection to glutamate and formate metabolism. Insertion of an additional A in the promoter region of the *ERG26* appeared from Ev02 to Ev16 but was not carried further. The *ERG26* gene codes for C-3 sterol dehydrogenase, which is involved in the removal of a C-4 methyl group from an intermediate of the ergosterol biosynthesis pathway. Mutants with loss of this gene are inviable[116]. Hence, the mutation in the promoter region might not be an inhibitor[116]. A point mutation leading to the insertion of a single nucleotide C (cytosine) in the Ori region between the VAR1 gene (codes for a mitochondrial 37S ribosomal protein) and mitochondrial 21S ribosomal RNA gene was identified from Ev06 cells, and this mutation continued to appear in the further stages of evolution. Bioinformatics analysis of this sequence using a webtool, Ori-Finder3 reconfirms that this mutation is present in ARS (Autonomously replicating sequence)

regions of the mitochondrial chromosome[117] (**Figure S 8**). A point mutation was observed in the intergenic region of the *COX3* and tRNA-Met in the Ev02 cell but it disappeared in later evolution rounds.

Table 5: List of mutation identified in the intergenic regions
Δ- deletion, e.g.: $A_{11\rightarrow12}$ – Example of a point mutation where additional A is added

Mutation	Ev02	Ev06	Ev16	Ev21	Annotation	Gene	Description
A→G	+	+	+	+	intergenic (-100/-807)	*GSH1* ← / → *LSB6*	glutamate--cysteine ligase/1-phosphatidylinositol 4-kinase LSB6
C→A	+	+	+	+	intergenic (-504/-184)	*MET6* ← / → *IES5*	5-methyltetrahydropteroyltriglutamate- homocysteine S-methyltransferase/Ies5p
$(C)_{5\rightarrow4}$	-	+	+	+	intergenic (+4733/-3179)	*VAR1* → / → *21S_RRNA*	mitochondrial 37S ribosomal protein VAR1/21S ribosomal RNA (Ori)
$(A)_{13\rightarrow14}$	+	+	+	-	intergenic (-240/+391)	*ERG26* ← / ← *EFM5*	sterol-4-alpha-carboxylate 3-dehydrogenase (decarboxylating)/protein-lysine N-methyltransferase
$(A)_{11\rightarrow12}$	+	-	-	-	intergenic (+264/-4749)	*COX3* → / → *tM(CAU)Q2*	cytochrome c oxidase subunit 3/tRNA-Met

5.4.3. Reverse engineering of *GDH1* mutation

Gdh1p is the main NADP-dependent enzyme involved in the synthesis of glutamate from ammonia and α-KG when yeast is growing on glucose[110]. In contrast, Gdh3p is an alternative isozyme that is suppressed in the exponential phase of the yeast growth but which becomes active for glutamate synthesis in the late exponential phase or when growing on non-fermentable carbon sources[109]. The genome sequencing data shows the *gdh1* mutation from Ev02. Additionally, PCR amplification and sequencing of the *gdh1* locus show the same mutation right from the Ev01. In other words, only those cells carrying *gdh1* mutation evolved to produce glycine and serine from formate in the evolution process. Thus, I re-engineered this mutation in the unevolved strain to check if the strain can be revolved to grow on formate and glucose. For this, I tested four clones of VBS18 (unevolved VBS10_*Δgdh1*) and two clones of unevolved VBS10 strains and compared their growth in SC-Ser-Gly medium supplemented with 100 mM $(NH_4)_2SO_4$, 100 mM glucose, 250 mM formate, incubated under a 10% CO_2 atmosphere.

The VBS18 clones start to grow after 25-30 days of incubation and reach saturation in 40 days (**Figure 16**). In contrast, the VBS10 clones did not grow. As expected, no growth is observed from both strain variants until the entire incubation period without formate supplementation (**Figure 16**). This result suggests the reversed engineered *gdh1* mutation is somehow helping the strains to operate the RGP in order to produce glycine and serine from formate in the tested conditions. Thus, it is possible to reproduce the formate-dependent growth in the VBS18, the unevolved VBS10 strain reengineered with *gdh1* mutation reintroduced.

In order to understand how gdh1 mutation might be contributing to operating the RGP and formate assimilation, I have drawn up a hypothesis as follows – Gdh1p is the primary enzyme that produces glutamate from ammonia and α-KG at the expense of an NADPH when yeast

is growing on glucose. For the operation of RGP, glucose (100 mM) and 100 mM of $(NH_4)_2SO_4$ were provided. High $(NH_4)_2SO_4$ concentration in the medium might provide higher driving forces for the intracellular Gdh1p activity to enhance glutamate synthesis, which is the primary intracellular organic nitrogen source and is generally found at high intracellular concentrations. Due to this, intracellular NADPH might be consumed more than usual, causing scarcity for other NADPH-dependent biochemical reactions. Thus, deletion of this gene might result in higher intracellular NADPH availability facilitating global NADPH-dependent reactions. The CH_2-THF producing reactions catalyzed by Ade3p (cytoplasmic) and Mis1p (mitochondrial) are NADPH-consuming reactions.

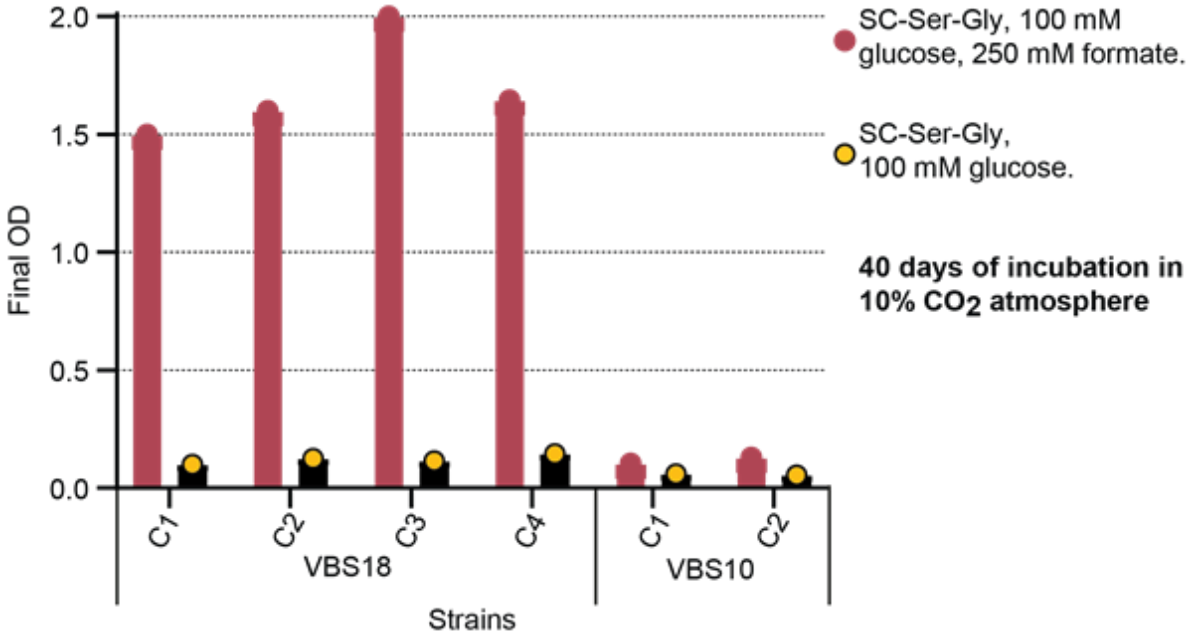


Figure 16: Re-evolution of reverse-engineered VBS10 strains
Final ODs of **VBS18 (VBS10_*gdh1:: Δ108bp*)** isolates (C1-C4) compared to VBS10 isolates (C1-C2) without *GDH1* deletion in SC-Ser-Gly medium with 100 mM glucose and 250 mM formate. VBS18 (C1-C4) started growing in 27 days and reached the indicated final OD_{600} in 40 days but was not grown without formate supplementation. **VBS10** (C1-C2) strains without *GDH1* mutation did not grow growth in both media conditions. **SC-Ser-Ger** = 1x YNB, 1x DO-Ser-Gly, 100 mM $(NH_4)_2SO_4$; **YNB** - yeast nitrogen base; **DO-Ser-Gly** - a mix of all the amino acids without serine and glycine.

To check if gdh1 deletion could influence formate assimilation for C1-THFs synthesis, I introduced the same gdh1 mutation in the S288c_WT strain and compared the CH_2-THF assimilation in S288c_WT and S288c_*Δgdh1* (VBS19 in Table 2) strains. A ^{13}C-labeling experiment was conducted to examine formate assimilation into histidine, serine & alanine, and methionine through CHO-THF, CH_2-THF, and CH_3-THF, respectively, in the S288c_WT and S288c_Δgdh1 strains. Biomass of S288c_WT and S288c_*Δgdh1* strains, fed with glucose and ^{13}C-formate in the SM medium and incubated in a 10% CO_2 atmosphere, were analyzed and compared for the percentage of histidine, methionine, serine, alanine, and glycine carrying

single or double-labeled carbons. Data analysis shows no significant differences in the labeling of histidine between both strains, indicating that CHO-THF synthesis is not influenced (**Figure 17**) by the gdh1 mutation. The CH_2-THF produced from formate is further reduced to CH_3-THF and assimilated for methionine synthesis, transferring a single ^{13}C carbon from formate to methionine. Thus, in the case of methionine, a significantly higher percentage of single labeled methionine is observed in S288c_*Δgdh1* compared to the S288c_WT strain. Similarly, a significantly higher single labeled serine is observed in the S288c_gdh1 mutant due to the direct CH_2-THF assimilation into serine (**Figure 17**). Additionally, 10% of alanine carrying in single labeled carbon from formate is observed in the mutant strain that is produced via deamination and transamination of single labeled serine, whereas the WT strain shows only 5% of single labeled alanine (**Figure 17**). These results suggest a clear improvement in the CH_2-THF synthesis where NADPH is used as a cofactor but not in ATP-consuming CHO-THF synthesis reaction. These labeling experiments confirm an increased synthesis of CH_2-THF from formate, leading to its enhanced assimilation in the S288c_*Δgdh1* strain compared with the S288c_WT control.

Approximately a little over 10% of single-labeled glycine and 6% of double-labeled serine were quantified in S288c_*gdh1:: Δ108bp*, which is considerably more than in the WT (**Figure 17**). I speculate that the single-labeled glycine and double-labeled serine are produced from the reversal of the native GCV system and Shm1p in mitochondria without any selection or overexpression of pathway modules. The percentages of single-labeled and double-labeled glycine and serine were observed only when a high concentration of $(NH_4)_2SO_4$ (100 mM) was provided in the medium but not when 40 mM of $(NH_4)_2SO_4$ was provided. This indicates that this minuet flux of single and double-labeled glycine and serine might be from the activity of native GSC.

As there is no selection for glycine/serine synthesis in these two strains, other possibilities of single labeled glycine synthesis cannot be ruled out, but as explained, this glycine synthesis was alerted in accordance with $(NH_4)_2SO_4$ in the medium. Thus, it is highly possible that the single labeled glycine might be via native GSC, as the growth of the sulfate-reducing bacteria that is growing via the RGP was reported to be influenced by $(NH_4)_2SO_4$ concentration provided in the medium[56,118]. However, a less single labeled alanine was observed when high $(NH_4)_2SO_4$ was provided in the medium as deamination of serine (Cha1p) to pyruvate might have a negative effect due to high intracellular NH_3. As reported, Shm1p is known to produce serine when glycine and CH_2-THFs are available; however, its overall activity is low compared to Shm2p[100].

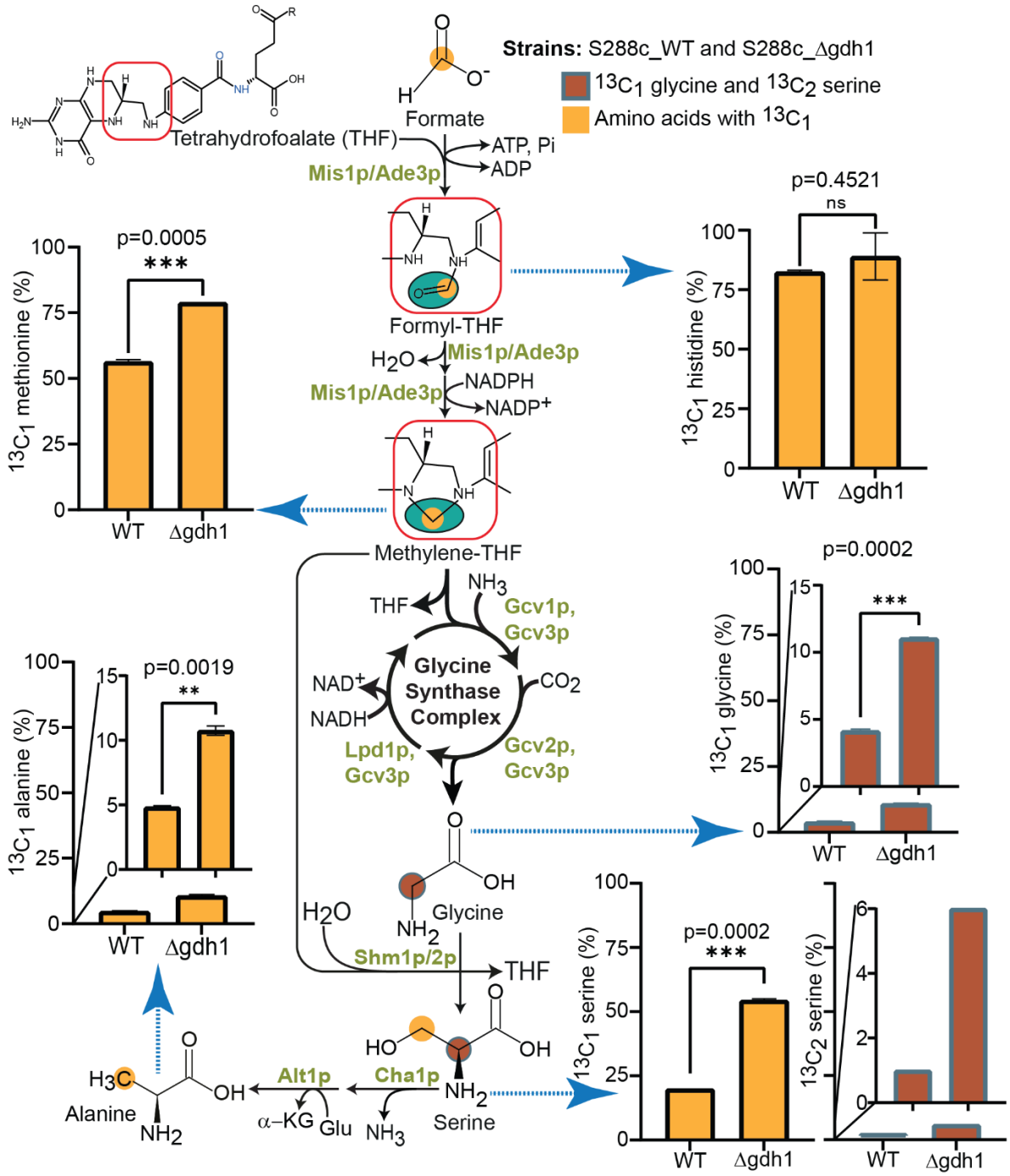


Figure 17: Comparison of formate assimilation in S288c_WT and S288c_*gdh1::Δ108bp* strains
Metabolic scheme of natural formate assimilation in routes in *S. cerevisiae*, ^{13}C labeled carbon flux is indicated via these pathways. Labeling percentages of the given amino acids from the biomass of S288c_WT and S288c_Δgdh1 grown on glucose and ^{13}C formate. The proteinogenic histidine, methionine, serine, alanine, and glycine carry single or double ^{13}C carbons. Statistical significance calculated between both strains by t-test, compared variance by F-test and p-value indicated. Statistics were not valid for $^{13}C_2$ serine. **Note:** *gdh1::Δ108bp* represent as Δgdh1 in figure for space convenience.

The results show improved cytoplasmic CH_2-THF synthesis and assimilation into methionine and serine. The mitochondrial and cytoplasmic pools of NADPH and C1-THF are discrete.

Thus, a precise mechanism is not established on how increased cytosolic NADPH pools and redox balancing mechanisms might alter its mitochondrial pool to show an effect on mitochondrial CH_2-THF synthesis that might lead to glycine and serine synthesis via the native GSC.

5.5. Pyruvate synthesis from formate

Glycine and serine synthesis from formate were established successfully in the VBS10 via the activity of the complete serine synthesis module of the RGP, ALE, and validated via the carbon tracing experiments (sections from **5.1.4 to 5.1.7**). In order to establish the complete pathway, the next step is to establish the third module of the RGP. The activity of the third module of RGP is to convert serine produced from formate in the RGP to pyruvate. This reaction is catalyzed by Serine deaminase, which is encoded by the *CHA1* gene in *S. cerevisiae*. In this reaction, serine is deaminated to pyruvate to connect formate with central metabolism leading to biomass production. Serine that is produced from formate carries two carbons from formate; when it is deaminated to pyruvate, both carbons are transferred to pyruvate, which is then trans-aminated to alanine and passes the same carbons from formate-derived serine to alanine (**Figure 19A**). Therefore, a double-labeled pool of alanine is an indicator for the pool of formate-derived pyruvate.

The first hint of pyruvate synthesis from formate via the RGP was observed in the labeling experiment, in which the proteinogenic amino acid analysis of the biomass of VBS10_Ev11 revealed approximately 10% of alanine and 4% of valine carrying two ^{13}C carbons (**Figure S 9A-B**). This indicates that serine produced from formate is deaminated to pyruvate, which is further transaminated to alanine. This is an indication of a pool of pyruvate synthesized from formate without any selection for its synthesis via the RGP. A significant flux of pyruvate comes from 100 mM of glucose in the medium via natural glycolytic flux. 250 mM formate that was provided in the medium used to produce all C1- units, glycine, serine, and a small pool of pyruvate/alanine by the evolved VBS10. Thus, I continued for a new evolution with Ev21 strains (refer **5.1.6. Figure 7**) while decreasing the glucose concentration in the medium to check if the selection pressure can be shifted from glucose to formate for pyruvate production by improving pyruvate/alanine flux via the RGP.

5.5.1. Evolution in glucose limiting condition

The new round of ALE was performed while gradually decreasing the glucose concentration in the medium with the Ev21 strains (**Figure 18**). In Ev22, glucose concentration was decreased to 40 mM, and no growth inhibition was observed. In subsequent evolution rounds, glucose concentration decreased to 20 mM in Ev25 and Ev26 and to 10 mM glucose in Ev27.

At these concentrations of glucose, growth inhibition started to appear, and a decrease in the biomass (final OD) was observed. Finally, glucose concentration was decreased to 5 mM and 2 mM, respectively, in EV28 and Ev29, and growth is further decreased to ~0.9 OD in Ev28 and 0.4 OD in Ev29. Even though growth is observed at glucose concentrations as low as 2 mM, biomass is reduced severely (**Figure 18**).

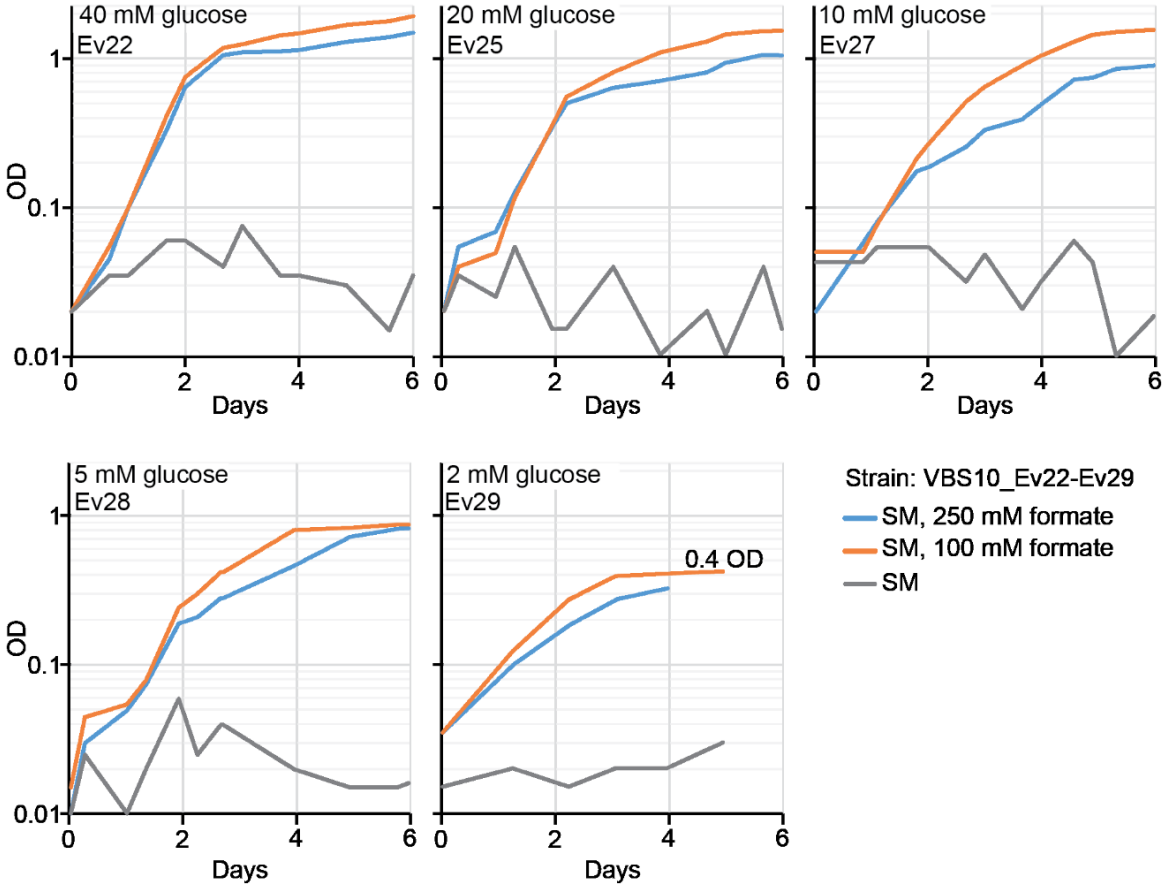


Figure 18: VBS10 strain evolution in glucose limiting conditions
The VBS10_Ev21 cultured in SM medium while decreasing the glucose conditions to drive more pyruvate flux from formate. Growth curves of the evolved VBS10 strains form Ev22, Ev25, Ev27, Ev28, and Ev29 in 40 mM, 20 mM, 10 mM, 5 mM and 2 mM glucose, respectively. **SM**= 1x YNB, 100 and 250 mM formate.

5.5.2. Labelling experiments to validate pyruvate biosynthesis

The evolved VBS10 strains from selected evolution rounds in the whole ALE process were assessed for ^{13}C labeling in alanine from ^{13}C formate to analyze for improvement in pyruvate flux via the RGP. Single colonies from Ev02, Ev06, Ev10, Ev16, Ev21, Ev25, and Ev29 were grown in SM medium supplemented with ^{13}C formate or/and ^{12}C formate and glucose. Biomass hydrolysates of selected strains were analyzed for the double-label alanine and found that approximately 15% -10% of alanine carries two ^{13}C carbons in the biomasses of the

evolved VBS10 strain. The WT shows no double-labeled alanine but less than 5% single-labeled alanine (**Figure 19B**). No labeled carbon from the biomass was observed when ^{12}C formate is used.

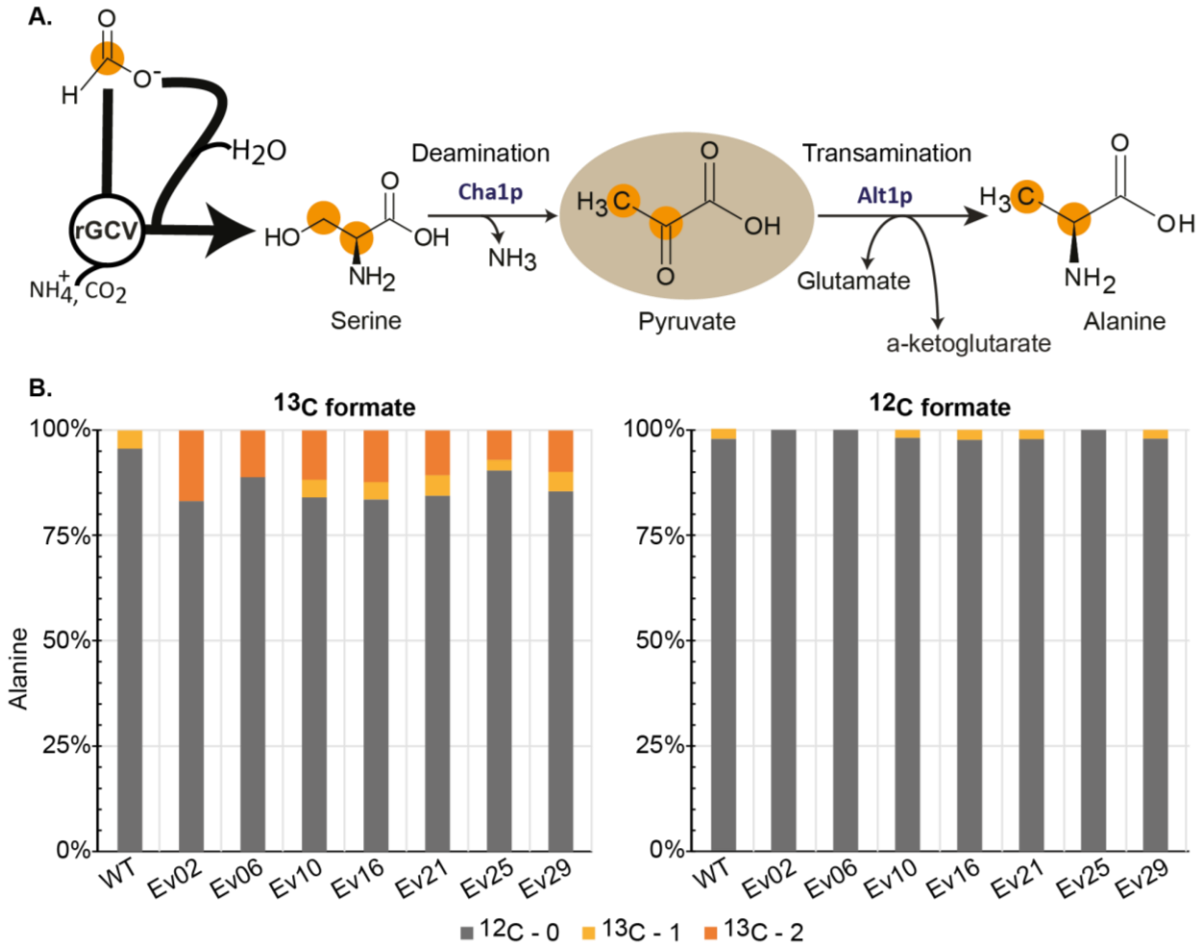


Figure 19: Confirmation of serine deamination in the evolved VBS10 strain
A. Metabolic scheme of pyruvate and alanine synthesis from formate via the RGP. Deamination and transamination of labeled serine produced from formate via the RGP. **B.** The double-labeled and unlabeled alanine ratio in evolved VBS10 when fed with $^{13}C/^{12}C$-formate and glucose in SM medium. A 15-10% of alanine was labeled twice in the evolved VBS10 strains but not in the WT strain.

The strain from Ev02 shows ~17% of $^{13}C_2$ alanine. The glucose-limiting evolution process did not improve serine deamination, as labeling of alanine was even slightly reduced in Ev25 and Ev29 compared to the early evolution rounds. Nevertheless, this result validates the activity of the third module of the RGP, i.e., serine deamination for pyruvate synthesis (**Figure 19A-B**) already in early evolution rounds.

These results confirm that serine produced via the RGP is deaminated by Cha1p, resulting in pyruvate synthesis, which is further trans-aminated to alanine by Alt1p (**Figure 19A**). To test if this additional pool of pyruvate from formate is resulting in more biomass in the VBS10 strain

compared to the WT strain, another growth experiment was conducted in limiting glucose conditions. In this experiment, VBS10_Ev26 and WT strains were fed with 20 mM of glucose and 100 mM of formate as the 20 mM glucose. The VBS10_Ev26 strain reached slightly higher biomass, ~1.8 OD_{600}, compared to the S288c_WT strain that reached only ~1.5 OD_{600} (**Figure S 11**). ^{13}C-formate labeling also validated the alanine/pyruvate synthesis from formate in this experiment (**Figure S 9B**). Even though this growth difference is less it was observed across different concentrations of glucose below 20 mM.

However, these strains did not grow when formate alone was provided as a sole carbon and energy source. Thus, further evolution was continued in the SM medium supplied with 2 mM glucose and 100/250 mM formate. The VBS10_Ev29 strains were repeatedly subcultured from Ev30 to Ev50, but no growth improvement was observed, and ODs did not rise beyond 0.4. This indicates that the strain did not evolve to improve flux to pyruvate (**Figure S 10**). Further plasmid transformations and genome manipulation were not successful in the evolved VBS10 strains. As a result, efforts to provide engineering assistance to further evolution failed.

5.6. Growth establishment of S288c on formate and serine

So far, the complete serine synthesis module of the RGP was engineered and established complete glycine and serine synthesis from formate. Pyruvate synthesis was also partially achieved via the activity of the native serine deaminase from the formate-derived serine, validating the complete RGP activity except for the energy module. However, full growth on formate as the carbon and energy source is not achieved, and further engineering and evolution efforts to improve the pyruvate flux from formate-derived serine to achieve complete formatotrophic growth were not successful so far.

As a next step, I decided to establish efficient conversion of serine to pyruvate by engineering the serine deaminase to establish growth using serine as the sole carbon source. Previous attempts to achieve consistent growth on serine as the sole carbon/energy source failed even with the overexpression of the *CHA1* (results not shown), which might be due to the degradation of serine via native SHMTs for energy and C1-THF production. Thus, I engineered the energy module (*FDH1*) to support the serine deamination module (*CHA1*) to establish the complete growth on serine and formate as carbon and energy source in the SM medium. The codon-optimized *FDH1* from *Candida boidinii* was employed for formate oxidation to produce NADH for energy support. To localize the NADH and CO_2 production to the mitochondrial compartment, a leader sequence, *PreCoxIV*, was prefixed to the *CbFDH1* coding sequence to localize ScFdh1p to mitochondria. The *PreCoxIV*, the leader sequence of cytochrome oxidase subunit IV, is shown to import proteins to mitochondria; here, I have shown in the

microscopic picture that the GFP is prefixed with *PreCoxIV* localized to mitochondria in yeast (**Figure S 1**).

The VBS20 was created on the S288c_WT background by transforming a pVB12 plasmid carrying *CHA1* and *mtCbFDH1* overexpression cassettes. A short-term adaptation of the VBS20 strain in the SM medium was conducted in two conditions 1. Serine as the sole carbon and energy source, and 2. Serine and formate are the carbon and energy. This short ALE yielded a strain capable of consistently growing on serine and formate (**Figure 20A; Ev01-Ev05**).

When serine and formate are both provided in the medium, formate is oxidized to CO_2 and NADH, which can provide reducing power to cellular needs of reducing power and energy, and serine can be deaminated to pyruvate and assimilated for biomass production. Hence, better growth is observed when formate and serine both were provided in the medium (**Figure 20B**). Whereas when only serine is provided in the medium, serine is oxidized to glycine, which is further degraded to formate and CO_2 to produce energy. As a consequence, only a very limited amount of serine might be available for pyruvate synthesis, thus leading to reduced ODs, and using a high concentration of serine can be toxic for cells. However, no growth was observed when formate was provided as the sole carbon and energy source (**Figure 20B**).

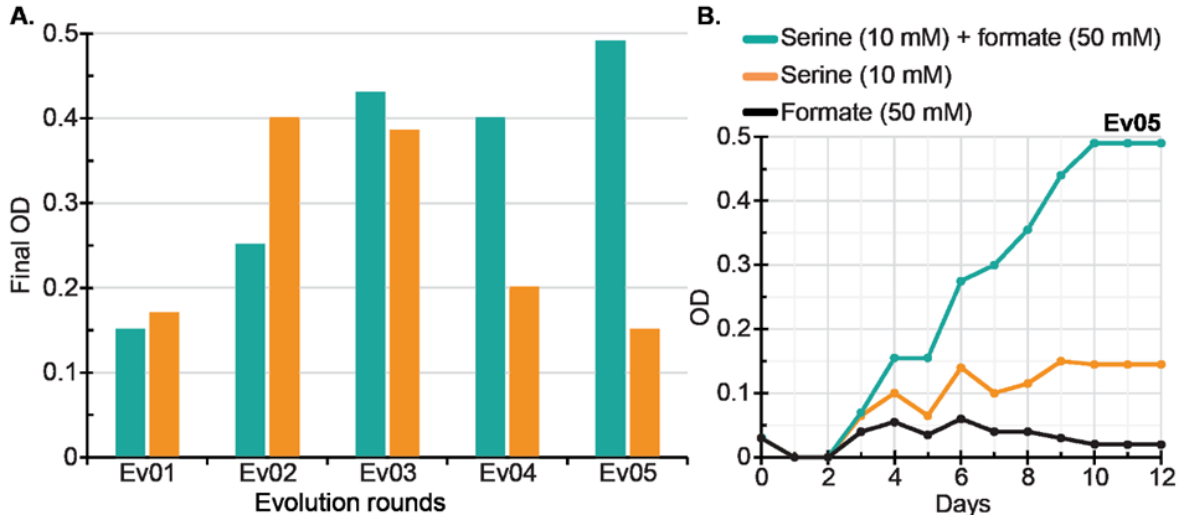


Figure 20: Short-term ALE of VBS20 using serine and formate as the carbon sources

A. Short-term ALE of VBS20 strain, carrying energy (*mtCbFDH1ox*) and serine deamination (*CHA1ox*) modules on pVB12 plasmid, in SM medium supplied with serine as the sole carbon source with or without formate supplementation as the additional carbon or/and energy source. Final OD_{600} improved from 0.15 to ~0.48 from Ev01 to Ev05. **B.** Growth curve of VBS20 in 10 mM serine with or without 50 mM formate, and 50 mM formate as the sole carbon and energy sources. **SM**= 1x YNB, 40 mM $(NH_4)_2SO_4$, 10 mM serine and 50 mM formate.

In conclusion, module 3 of the RGP, with the support of the energy module (module 4), supports full growth of the VBS20 strain on SM medium supplemented with serine and formate as the carbon sources. Further, this strain can be engineered with a complete pathway and can be evolved under different concentrations of formate while reducing serine in the medium to establish full growth of *S. cerevisiae* using formate as the sole carbon and energy source.

6. Discussion

In this study, I demonstrated the activity of all three of the four modules of the RGP together, i.e., glycine, serine, and pyruvate synthesis modules that connect formate to pyruvate in ΔS strains, except the energy module involved in formate oxidation to CO_2 and NADH for the energy support. I established complete glycine and serine synthesis from formate and CO_2 and demonstrated the native serine deaminase (Cha1p) activity, which led to pyruvate (alanine) synthesis. Further, I established the complete growth on serine and formate as the carbon and energy sources in the S288c_WT background overexpressing native *CHA1* as the pyruvate synthesis module, and *mtCbFDH1*, as the supporting energy module. Further, these results are discussed in the following sub-sections.

6.1. Serine biosensor strains and differences

I employed two prototrophic *S. cerevisiae* strains, i.e., S288c and FL100, to construct ΔS strains and established the RGP in these strains. The S288c variant of base ΔS strain (VBS01) carries Δ*agx1*, Δ*gly1*, Δ*ser1* knockouts which insulate serine, glycine, and C1 metabolism (**Figure 4**). Serine and glycine interconversion remains naturally active, meaning serine degradation to glycine and CH_2-THF and vice versa is naturally carried out by Shm1p in the mitochondria and by Shm2p in the cytoplasm (**Figure 4**)[100]. In the FL100 variant, two different ΔS strains were created that are VBS13 and VBS14. Both these strains carry Δ*agx1*, Δ*gly1*, and Δ*ser1* like VBS01 (S288c background) (**Figure 4**); additionally, VBS13 has Δ*shm1* (Figure S4A), and VBS14 carries Δ*shm2* (**Figure S 4B**). The Δ*shm1* mutation disrupts serine synthesis reaction in mitochondria, and the Δ*shm2* blocks the same reaction in the cytoplasm. Hence, glycine conversion to serine or its reversal is confined to the cytoplasm in VBS13, but in contrast, the same is restricted to the mitochondrial compartment in the VBS14 strain (**Figure S 4A & 4B**) (please refer table 2 for the information on strains).

Different phenotypes of these ΔS strains (VBS01, VBS13, and VBS14) from separate experiments were observed in response to the supplementation of glycine, serine, and formate (glucose is always present as the primary carbon source in all the above carbon source combinations). The VBS01 grew when glycine was supplied (**Figure 5B**); in contrast, both FL100 variants (VBS13 & 14) did not grow with glycine supply (**Figure S 4C & 4D**). Yet, when glycine and formate were provided, the VBS01 strain did not grow (Figure 6C), and a contrasting phenotype was observed in the case of VBS13 and 14 as both strains grew well when formate and glycine were supplied (**Figure S 4C & 4D**). In order to establish growth in the presence of formate and glycine in the base S288c_ΔS strain, VBS04, VBS05, and VBS06 strains were created by genome integration of *SHM1*, *SHM2*, and *EcGlyA* overexpression

modules into VBS01, respectively. Overexpression of SHMTs allowed glycine conversion to serine in the presence of formate in S288c_ΔS variants (**Figure 6**), which is essential for the RGP activity. The S288c_ΔS strain carrying EcGlyAp (VBS06) showed better growth than the other variants, indicating the better activity of EcGlyAp compared to *S. cerevisiae* isozymes, i.e., Shm1p and Shm2p.

Additionally, as discussed below, other differences were found between VBS13 and VBS14. Serine supplementation to the medium rescued growth of VBS13 (Δ*shm1*) but not VBS14 (Δ*shm2*). According to the previous report, cytoplasmic Shm2p contributes 95% of total SHMT activity, and Shm1p in the mitochondrial compartment carries out the rest. Especially in the serine supplied medium, Shm2p is the primary isozyme that performs serine degradation and supplies glycine and C1-moieties to the central metabolic network. In this condition, Shm1p is known to function in the reverse direction utilizing glycine by GCV to produce CH_2-THF and convert glycine to serine in the mitochondrial compartment[100,105,106]. Thus, serine addition rescued the growth in the FL100_ΔS strain with Δ*shm1* (VBS13). In the Δ*shm2* (VBS14) variant of FL100_ΔS, no growth was observed when serine was provided to the medium; this indicates that there might not be serine degradation by Shm1p, or at the best not enough to support growth as the overall activity of the Shm1p is deficient. This result complies with previous reports that Shm2p is the major supplier of C1-units when yeast grows on serine as the source C1-units, and when glycine is provided, Shm1p isozyme contributes to serine synthesis[100,106].

Both of the FL100 variants (VBS13 & 14) failed to grow when formate was supplemented along with serine (**Figure S 4C & 4D**); these results contrast with the VBS01 (S288c_ΔS variant) results where the growth of this serine biosensor strain is rescued when serine and formate are supplemented (**Figure 5C**). In the case of the FL100_ΔS strains, it can be speculated that addition of formate might have generated an adequate number of C1-units to donate the C1-moiety for other biochemical reactions; thus, *SHM1/2*, which are responsible for the generation of intracellular C1-units when serine is present, might be repressed. According to the regulation of one-carbon metabolism, a model has been proposed in which the availability of CH_2-THF inhibits serine degradation by cytoplasmic SHMT[106]. However, a different phenotype was observed in the S288c variant (VBS01), here growth was found when serine and formate were provided in the medium (**Figure 5**). In conclusion, several phenotypic differences were observed among the ΔS strains constructed in the backgrounds of S288c *MATα* and FL100 *MATa*. In conclusion, glycine to serine conversion in the S288c_ΔS strains was enabled via the overexpression of one of the SHMTs and whereas FL100_ΔS strains can perform the same reactions without overexpression of *SHMTs* in the RGP operable conditions.

6.2. ALE and differences across serine biosensor strains

The ΔS strains constructed under the S288c and FL100 background, expressing the complete serine synthesis module or only the glycine synthesis module, were employed for evolution to establish glycine and serine synthesis from formate in a high CO_2 atmosphere. The VBS08, VBS09, and VBS10 are S288c variants overexpressing the complete serine synthesis module of the RGP, and the VBS15 and VBS16 strains are FL100 variants expressing only glycine synthesis module; thus, these latter strains rely on native Shm1p/Shm2p activity for the glycine to serine conversion. All of these strains were employed in the ALE process to establish glycine and serine synthesis from formate via the RGP.

The VBS10 strain, overexpressing *EcSHMT*, evolved in a semi-rich medium (SC-Ser-Gly) and showed successful glycine and serine synthesis from formate and CO_2 (**Figure 7**). Later, this strain was shifted from semi-rich medium to minimal and ALE further improved its formate and CO_2-dependent growth (**Figure 8**).

The VBS08 and VBS09 overexpressing *SHM1* and *SHM2* failed to show any growth in the identical semi-rich medium, hence failing to produce glycine and serine from formate and CO_2 (**Table 3**). Thus, a different ALE process was adapted to evolve VBS08/09 in glycine limiting conditions, as explained in subsection **5.1.9.** However, VBS08 broke the selection of serine biosensor strain and evolved wrongly to produce glycine and serine from glucose (**Figure 10D & Figure S 3A**). VBS09 did not survive the repeated culturing in lower glycine (0.05 mM) conditions (**Figure S 3B**) and thus failed to evolve.

The FL100 variants, VBS15, and VBS16 strains were directly cultured in minimal (SM) medium. The VBS15 strain with cytoplasmic glycine to serine conversion did not grow in the SM medium supplied with formate and glucose though it was incubated in a 10% CO_2 atmosphere for a very long period (>45 days). In contrast, a 45-day prolonged incubation of VBS16 in the same conditions resulted in growth, synthesizing glycine and serine from formate and CO_2, which was further improved via the ALE process (**Figure 11A & 11B**). In this strain, the glycine conversion to serine is confined to mitochondria and is carried out by native mitochondrial Shm1p (**Figure S 4B**). Even though both native *SHM1/2* are present in VBS08 (of S288c variant), glycine to serine conversion is expected to rely on *SHM1* overexpression (**Table 2**), in which the selection was breached (**Figure 10D**). The native Shm1p activity accompanied by activity of the overexpressed glycine synthesis module successfully supported formate and CO_2 dependent growth in the FL100 variant VBS16 but not in the S288c variant that is VBS08 in which whole serine synthesis pathway is overexpressed (**Figure 11A-C**). However, these two strains, VBS08 and VBS16, are not identical genotypes.

The VBS09 strain carries an overexpression module of *SHM2*, and the VBS15 relies on native *SHM2* for glycine to serine conversion. In both cases, glycine to serine conversion is largely confined to the cytoplasm; both these strains did not grow when glycine was withdrawn/limited from the SM medium and thus, did not evolve. It is known that limiting glycine in the medium leads to transcriptional repression of *SHM2*[119]. However, VBS09 also failed to evolve even though it carries an overexpression cassette of the *SHM2* module, which should not be repressed when glycine is limited in the medium. Additionally, the VBS10 strain expressing *EcSHMT*, which is also localized to cytoplasm, evolved successfully. This indicates that it is not the location in the cytosol *per se* or the expression pattern, but some unknown feature of the SHMT protein that prevents it supporting glycine to serine conversion. Previous reports also says Shm2p might have additional roles in other non-catalytic cellular function apart from its SHMT activity[120].

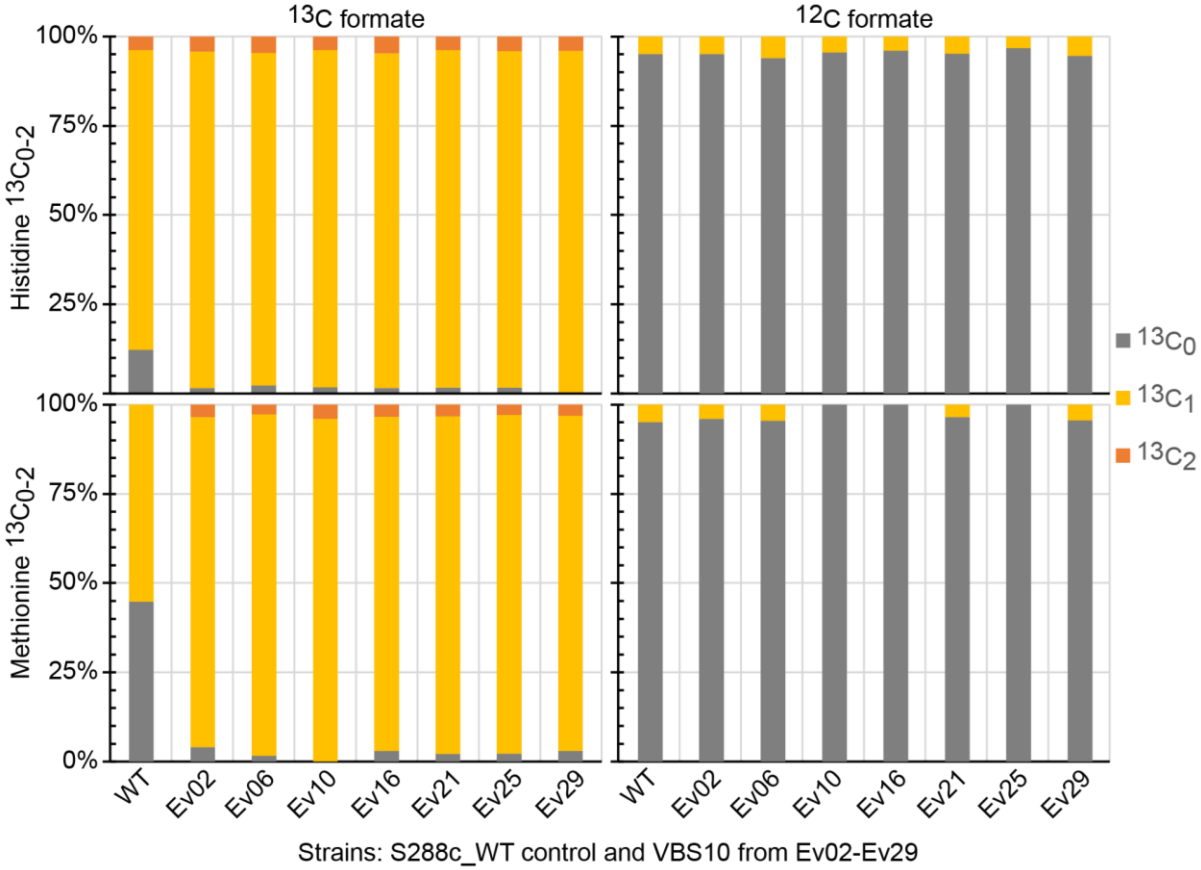


Figure 21: Comparison of C1 unit synthesis between WT and evolved ΔS strains
Carbon tracing experiments with ^{13}C formate show almost complete C1 unit are synthesized from formate that is assimilated into histidine and methionine in the evolved biosensor strain, whereas in the WT strain, relatively fewer C1-unites are generated when formate is available without insulation of C1 metabolism. This infers that C1 unit synthesis pathways from formate are naturally active in yeast.

Summarizing, among the five different serine biosensor strains, i.e., VBS08, VBS09, VBS10, VBS15, and VBS16 employed in the ALE, only VBS10, and VBS16 successfully evolved for

C1 unit (**Figure 21**), glycine and serine synthesis from formate and CO_2 via the RGP. (Note: C1 units generated naturally assimilated in to histidine and methionine (**Figure 21**)). However, my experiments did not establish a conclusive explanation for why the VBS08, VBS09, and VBS15 strains failed to produce biomass via the RGP using formate and CO_2 as the additional carbon sources and glucose as the primary carbon source.

6.3. RGP dependent growth and bottle-necks

The RGP operates by assimilating CO_2, formate, and ammonia to produce glycine, serine, and pyruvate. The evolved strain (VBS10_Ev21) that produce RGP-dependent biomass by assimilating formate and CO_2 was tested for their dependency on the three substrates, i.e., formate, ammonium sulfate, and CO_2. Different substrate concentrations were tested to find the suitable concentration of each substrate that drives the RGP optimally, using growth as a readout in the evolved VBS10 strain. The highest and non-deleterious concentrations of all three substrates were provided in the medium during the evolution process to provide better driving forces for the pathway.

The natural flux of the GCV system is the oxidation of glycine in ambient CO_2 conditions releasing CO_2, NH_3, and CH_2-THF. The reversal flux of GCV leads to glycine synthesis via the so-called glycine synthase complex (GSC). In the first reaction, the Gcv1p assimilates the CH_2 group from CH_2-THF produced from formate, combines it with NH_3 derived from $(NH_4)_2SO_4$, and attaches this amino-methyl molecule to the carrier peptide of the GSC, Gcv3p.

When the effect of formate and $(NH_4)_2SO_4$ concentrations are tested on the GSC operating strains, the growth rates of the VBS10_Ev21 strain are optimum when the formate concentration is between 75-125 mM. The formate concentrations below 50 mM led to inconsistent growth, indicating that the RGP kinetics might be affected at these formate concentrations, whereas, above 300 mM format, growth inhibition occurs that is maybe due to toxicity (**Figure 12**). Even though a slight difference in the labeling of glycine concerning $(NH_4)_2SO_4$ concentrations was observed, the growth rates of the VBS10_Ev21 did not change when $(NH_4)_2SO_4$ concentration was changed between 10 mM and 120 mM (**Figure 13**).

The Gcv2p enzyme in the second reaction of GSC, a CO_2 molecule, is attached to the amino-methyl group on the Gcv3p, releasing a glycine molecule. The Gcv2p reaction requires high CO_2 conditions to assimilate the CO_2 molecule, which is opposite to the natural reaction leading to CO_2 release. Elevated levels of CO_2 over ambient CO_2 levels are needed to reverse the Gcv2p reaction. The experimental data on the requirement of CO_2 for the RGP-dependent growth suggest the same that the CO_2 concentration in the incubation atmosphere strongly affects the growth rates of evolved VBS10_Ev21 strains as the RGP is dependent on the high

CO_2 concentration for the reversal of the GCV system[54,55,61]. A 5% percent CO_2 atmosphere is optimum for the growth of the evolved VBS10_Ev21 strain (**Figure 14A-D**). An increase to 10% CO_2 did not improve the growth rate; thus, pathway activity/flux might not be enhanced with the increase in the CO_2 content over 5%. The VBS10_Ev21 strain grew even under a 2.5% CO_2 atmosphere, which is a 50% reduction in the high optimum CO_2 concentration, but the growth rate was reduced by more than a half compared to optimum growth conditions (**Figure 14A-D**). Despite the growth reduction, the RGP is operable in the 2.5% CO_2 condition.

The experimental data shows that 2.5% CO_2 can reverse the Gcv2p reaction but at a lower rate than 5% CO_2. Thus, using this data as a basis, the VBS10_Ev21 strain can be further evolved by continuously culturing under the 2.5% CO_2 atmosphere to improve the growth rates of these strains under the reduced percentages of high CO_2 to enhance glycine synthesis via the GSC. In a different approach, the CO_2 assimilation activity of the Gcv2p enzyme can be improved via the directed evolution of the enzyme using in vivo rapid mutagenesis or *Insilco* enzyme engineering methodologies[121,122].

6.4. Endogenous serine deaminase and restrictions on pyruvate formation

The carbon tracing experiments with ^{13}C formate/$^{13}CO_2$ validated that C1-units (**Figure 21**), glycine and serine are synthesized from formate via the complete serine synthesis module of the RGP in both S288c and FL100 variants of ΔS strains, i.e., VBS10 (**Figure 9B**) and VBS16 (**Figure 11C**). The selection strategy of ΔS strains offers selection pressure to test the pathway up to serine synthesis (The complete serine synthesis module). Thus, the ΔS strains were employed to test for glycine and serine synthesis from formate and CO_2 via the overexpression and evolution of the complete serine synthesis module of the RGP.

The next step in the RGP is to deaminase the serine that is produced from formate in order to produce pyruvate and thus to connect the RGP to the central metabolic network. However, the ΔS strain does not offer selection pressure for the pyruvate synthesis as pyruvate is freely produced from the glucose provided in the medium. Nonetheless, serine deamination occurred at a low rate, converting some formate-derived serine to pyruvate in the evolved ΔS biosensor strains. These pyruvate molecules are further trans-aminated to alanine which is assimilated for biomass production. The sequential serine deamination and alanine transaminase reactions transfer two ^{13}C carbons from serine to alanine. Thus, the alanine from the biomass carrying two ^{13}C carbons from formate is indicative of pyruvate flux via the RGP (**Figure 19A**).

Consistently, 10-17% dual labeled alanine was observed from the biomass of the evolved VBS10 strains across several evolution rounds, i.e., Ev02, Ev06, Ev10, Ev16, Ev21, Ev25, and Ev29 (**Figure 19B**). Double ^{13}C labeled alanine, double-labeled serine, and single ^{13}C labeled glycine validate the activity of all three major modules of the RGP that connect formate to the central metabolic network at pyruvate. Additionally, approximately 2.5 to 5% of dual labeled valine was observed in the evolved VBS10 strain, although this was not very consistent (data not shown). The VBS16 (FL100_ΔS) strain also showed traces of dual labeled alanine; however, this represented less than 5% of total alanine (data not shown). Nonetheless, the RGP was demonstrated in the evolved VBS10 strains, with the exception of the energy module. Though all three modules of the RGP are active, full growth on formate as sole carbon source was not achieved. I speculated that the serine deamination might be inefficient maybe due to expression constraints or to regulation.

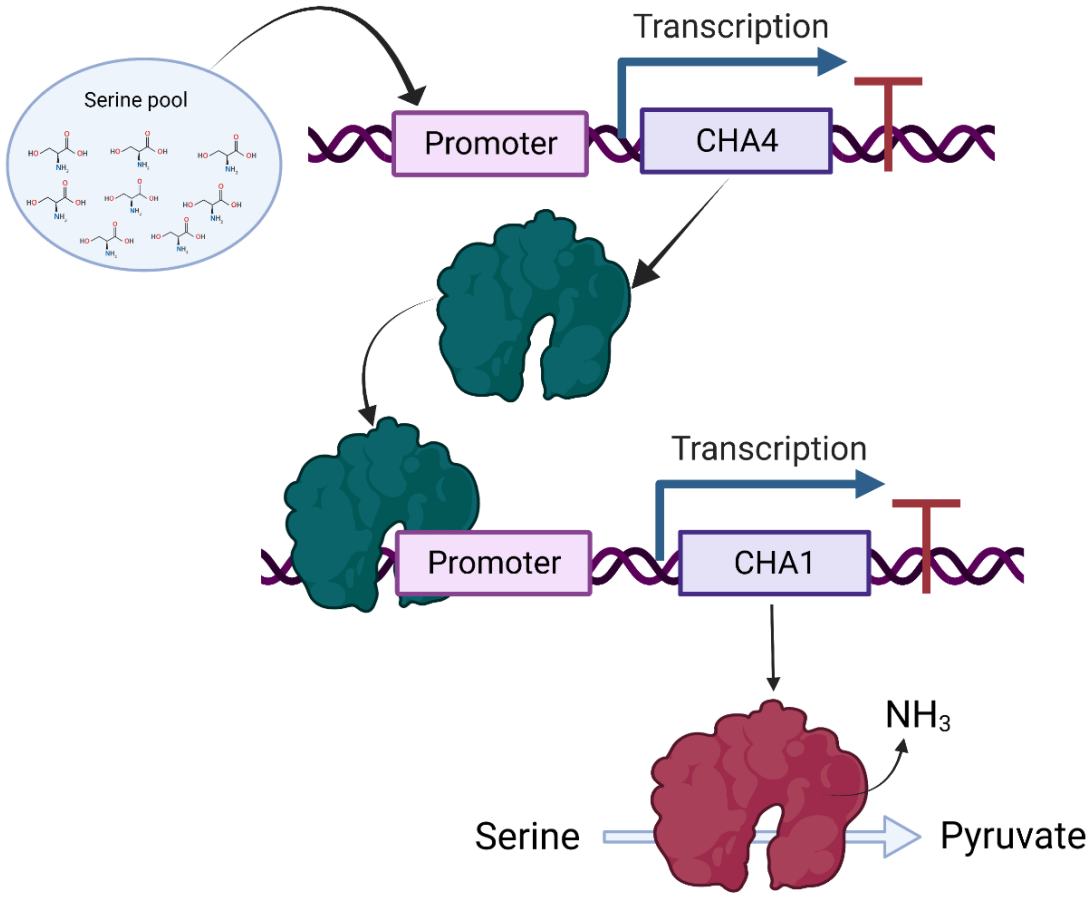


Figure 22: Mechanism of CHA1 expression leading to serine deamination
Accumulation of intracellular serine or threonine (hydroxyl amino acids) pool drives the expression of *CHA4*, a DNA-binding transcriptional activator that binds in the promoter region and initiates the expression of *CHA1*. The Cha1p (serine deaminase) converts available serine to pyruvate. Thus, the native serine deamination occurs in response to the serine pools, especially in the late exponential or early stationary phase. This figure was "Created with BioRender.com".

In yeast, serine deaminase (Cha1p) is encoded by the *CHA1* gene, and its expression is controlled by transcriptional activator Cha4p in response to the intracellular accumulation of hydroxy amino acids, i.e., serine and threonine. The accumulated serine induces the expression of *CHA1* mediated through the DNA binding transcriptional activator, Cha4p. Upon the expression in response to the serine accumulation, the Cha1p converts serine to pyruvate (**Figure 22**)[123–125].

Most of the glycine and serine flux from formate via the RGP might be actively assimilated for biomass production when cells are growing in the exponential phase. Nonetheless, serine accumulation seems to be a possibility between the late exponential phase and the early stationary phase. These serine pools might have been deaminated to pyruvate and transaminated to alanine. This might be leading to just 10%-17% of alanine produced from the formate derived-pyruvate via the RGP before cells enter into saturation. Formate-derived pyruvate might be very limited for re-assimilation into the TCA cycle to supply reducing power and energy for the RGP operation to maintain the intracellular serine pools on which *CHA4* expression is dependent. Lack of formate-derived serine pools and *CHA1* expression limit pyruvate synthesis, thus leading to minimal pyruvate flux via the RGP in the late exponential phase. Thus, an active pyruvate synthesis/serine deamination module is needed while the cells are growing in the exponential phase to improve the RGP based pyruvate flux via active deamination of formate-derived serine.

However, overexpression of *CHA1* or *CHA4* in order to establish the active serine deamination is a possible solution to increase conversion of serine to pyruvate in the exponential phase. However, this approach failed due to known difficulties of transformation to introduce the *CHA1* overexpression plasmid and genome manipulation in the evolved VBS10. Different transformation methods like glycerol and sorbitol-based electroporation along with traditional PEG-based heat shock transformation methods were tested, but failed. The evolved VBS10 cells from different evolution rounds did not survive the transformation procedure in many attempts, thus, it was not possible to engineer the evolved VBS10 strains further to establish efficient serine deamination.

However, the evolved VBS10 strains are ΔS strains and do not provide selection pressure for pyruvate synthesis; thus, primarily, pyruvate, alanine, and other metabolites are produced from glucose. Hence, in these strains, most cellular components are produced from glycolytic and OPP pathways. Only C1-units, glycine, and serine are produced from formate via the RGP, and non-assimilated serine at the end of the exponential phase might have been converted to pyruvate. Hence, I decided to decrease the glucose concentration in the medium to increase the pyruvate synthesis burden on the RGP. Evolution in glucose limiting conditions

was opted to proceed while decreasing the glucose concentration and expecting to use formate as the primary carbon source for the synthesis of RGP-dependent pyruvate leading to more complete biomass production from formate and CO_2 feedstocks.

6.5. Evolution efforts for complete formatotrophy

The VBS10_Ev21 strain was further evolved in the minimal medium while reducing the glucose concentration to enhance native serine deamination. The native serine deaminase is a non-essential gene (*CHA1*); functionally, it is localized to the mitochondria. Although the native Cha1p is functional, its expression is probably constrained in the exponential phase due to the lack of serine accumulation because of the continuous utilisation of serine for biomass production in exponentially growing cells and/or because of the low rate of serine synthesis from formate and CO_2. One aim of ALE in the glucose limiting condition was to adapt the evolved VBS10 strains for the low glucose conditions. Further aims were to test if the strains can self-optimize to such physiological conditions to produce energy for the operation of the RGP and to test if they can then maintain intracellular serine pools to operate the serine deamination to contribute for the enhanced biomass by improving pyruvate flux via the RGP from formate in high CO_2 conditions (explained in section **5.5.1**, **Figure 18** and **Figure S 10**).

The evolution in glucose-limiting conditions was continued up to 50 rounds (Ev50). From Ev22, glucose concentration was decreased to 2 mM until Ev29 while maintaining 100 or 250 mM formate in the medium. In this phase, growth decreased 2 OD_{600} from to 0.4 OD_{600} with a decrease in glucose concentration (**Figure 18**). Further evolution was continued up to Ev50 in 2 mM glucose condition to check if the strain can produce higher biomass than 0.4 OD_{600} utilizing formate. However, the evolution did not lead to improved biomass production. Biomass/growth decrease concerning decrease in glucose indicates that, still, glucose is the primary carbon and energy source. This phenotype also indicates the lack of NADH synthesis from formate oxidation to operate the RGP to maintain the intracellular serine pool.

The yeast *S. cerevisiae* is known to have two formate dehydrogenases, i.e., *FDH1* and *FDH2*. Activity Fdh1p/2p leads to increased biomass when formate is added as an auxiliary substrate in glucose-limiting conditions. A previous report demonstrated that Fdh1p activity did not lead to higher biomass in *S. cerevisiae* CEN.PK113-7D and Fdh2p activity leads to improved biomass production in CEN.PK 113-7D when formate is supplemented as auxiliary substrate[126]. This report showed that the deletion of *fdh1* did not lead to loss of formate dehydrogenase activity in this strain, but the addition of *fdh2* deletion on top resulted in the loss of formation consumption in the CEN.PK 113-7D[126]. This indicates that Fdh2p is the major contributor to formate oxidation and NADH synthesis. However, the S288c parent strain that

is used in this study has a truncated version of *FDH2*, with a stretch of 422 bp after the starting codon being lost (**Figure S 12**).

A complete coding sequence of *FDH1* is present in the S288c strain; however, no improvement in the growth of S288c was observed when formate was added to glucose-limited medium. Previous reports on formate dehydrogenase activity again in CEN.PK strain says that a considerable effect of NADH, synthesized via formate dehydrogenase activity, on the glycerol/ethanol synthesis pathway was only observed when *either of the* formate dehydrogenase was overexpressed either localized to cytoplasm or to mitochondria, not otherwise[127]. Up to this point in this study, neither *FDH1* nor *FDH2* is overexpressed; thus, a pronounced effect of NADH via the formate oxidation influencing biomass production was not observed in the S288c. Yet, a minimal formate dehydrogenase activity cannot be strictly excluded, even if it is not influencing growth. An additional indication of energy scarcity is when the evolved VBS10_Ev50 strains were grown on the agar plate with a minimal medium containing a 2 mM glucose and 100 or 250 mM formate, elongated cell growth was observed with pseudo-hyphal formation. The phenomenon indicates the energy limitation directed towards lack of energy/reducing power synthesis from formate.

Nevertheless, due to either lack of efficient energy (energy module) support or lack of efficient serine deamination (pyruvate module), or the collective effect of lack of both modules in the evolved VBS10 strains, further growth improvement was not achieved even at Ev50 (**Figure S 10**). Thus, this evolution for full biomass production from formation and CO_2 in glucose-limiting condition failed. Hence, I decided to overexpress *CHA1* in the S288c WT strain to establish growth on serine as the sole carbon/energy source for pyruvate synthesis via the serine deamination reaction to check if full growth can be achieved on serine with glucose.

6.6. Growth of S288c on serine and formate as the feedstocks

As a next step to establish yeast growth on serine, the third module of the RGP overexpressed, that is *CHA1*. The S288c_WT did not show considerable growth when serine was provided as the sole carbon/energy source, even with longer incubations. Considering that the native serine deaminase is active in response to intracellular serine, two scenarios can explain the no-growth phenotype of S288c_WT. Those are, either there could be a problem with serine uptake or serine entered into the cell might have been oxidized by native *SHMTs*, especially by the highly active cytoplasmic Shm2p, resulting in unavailability of the serine for pyruvate synthesis.

Nevertheless, the ΔS strain is able to take up serine provided in the medium as the ΔS strain grows upon the addition of serine to the medium with glucose as the primary carbon source.

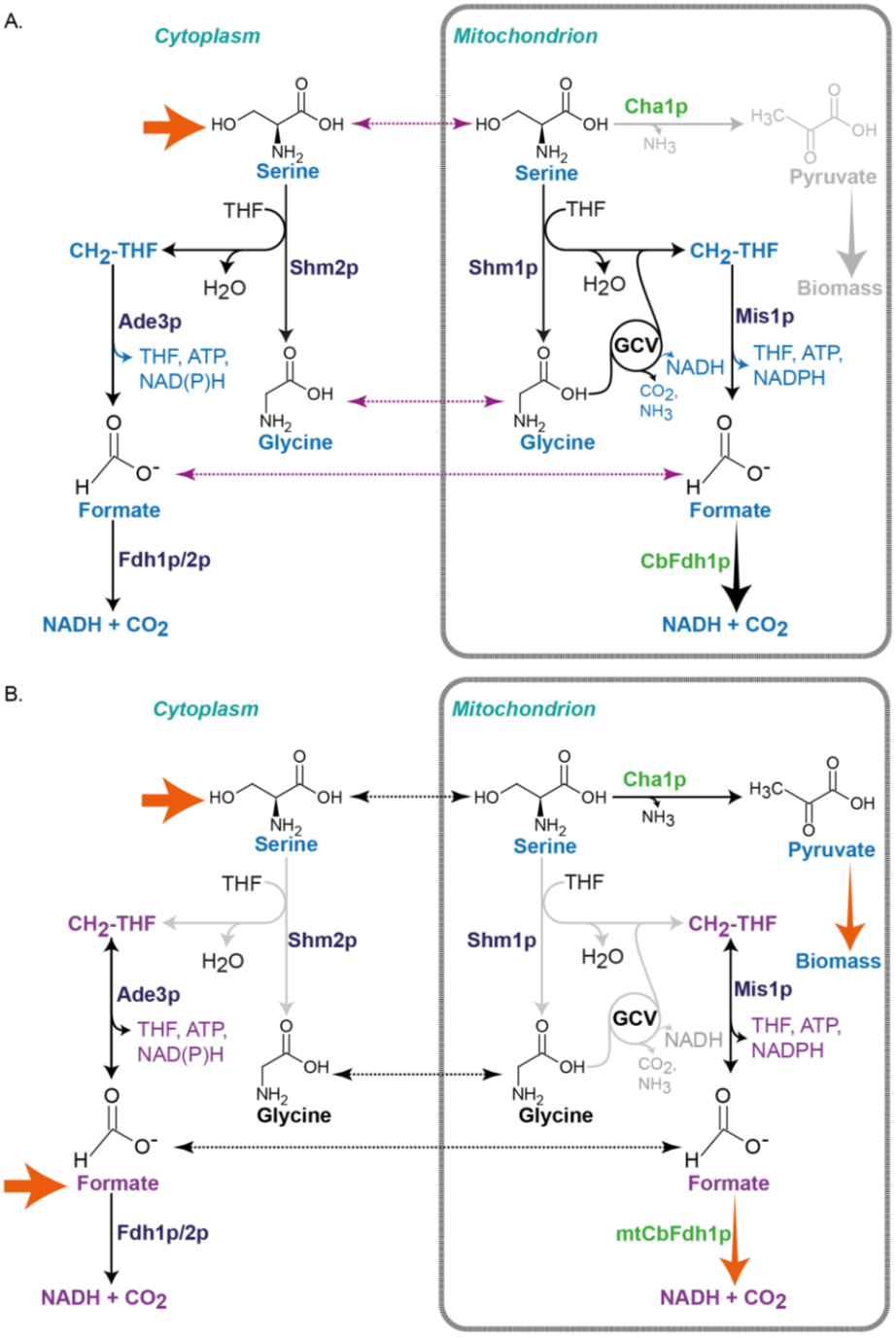


Figure 23: Metabolic schematic energy and biomass synthesis in the VBS20 strain

VBS20 strain expressing *ScCHA1* (serine deamination module) and *mtCbFDH1* (energy synthesis module) on the plasmid. **A.** When serine (blue color) is supplied as the sole carbon and energy source, serine is oxidized to glycine and CH_2-THF in the cytoplasm and mitochondria. Glycine is further oxidized in the mitochondria via the native GCV system to CO_2, NADH, and CH_2-THF. CH_2-THF is further oxidized to formate, releasing ATP and NADPH molecules in both the cytoplasm and mitochondria. The formate is further oxidized to CO_2 and NADH. Due to the consumption of serine for energy and reducing power, a limited amount of serine is available for pyruvate synthesis. **B.** When formate and serine are supplied (blue and purple colors), formate can partially be oxidized to the CO_2 and NADH and reduced to CHO-THF, CH_2-THF, and CH_3-THF. So that energy, reducing power, and C1-units synthesis can be covered from formate so that serine can be primarily used for pyruvate synthesis leading to better biomass production. (Light gray line: possible reduced metabolic flux)

Yet, when serine was provided as the sole carbon source, growth was not enabled in ΔS strain overexpressing *CHA1*. This indicates that serine uptake might not be a problem but that, rather, the native SHMT activity might be oxidizing serine to glycine and CH_2-THF that will restrict serine availability for pyruvate production[100].

Thus, I opted for a different approach to overexpress both pyruvate and energy synthesis modules together in the S288c background, which is VBS20 from the table of strain list (**Table 3**). The VBS20 strain carries a native *CHA1* overexpression module for efficient serine deamination. To establish an energy supply module, I overexpressed *FDH1* from *C. boidinii* with a leader sequence to target and localize the protein into mitochondria. The mtCbFdh1p-based energy module oxidizes formate into CO_2 and NADH to supply reducing power and energy for intracellular requirements. *C. boidinii* is one of the methylotrophic yeasts. Methylotrophic yeasts are known for their efficient formate dehydrogenase -based detoxification[128,60,129]; thus, codon optimized *FDH1* from *C. boidinii* was established in S288c along with serine deamination module, Cha1p functionally localized to the mitochondria.

A short-term ALE led to VBS20 growth on serine and formate as the carbon and energy sources and also to growth using serine alone as the carbon and energy source. However, stable growth is observed only when both serine and formate are provided in the medium for energy and biomass production (**Figure 20**). When serine is provided as the sole carbon source, serine can be oxidized to glycine and CH_2-THF via the native Shm1p and highly active Shm2p (**Figure 23A**). Glycine can be further oxidized via the native GCV system leading to NADH and CH_2-THF production in the mitochondria. CH_2-THF is further oxidized to formate in the cytoplasm and mitochondrial via the Ade3p and Mis1, respectively, which leads to ATP, and NADPH synthesis. Formate is further oxidized to CO_2 and NADH synthesis. Consumption of serine in this oxidative metabolism curbs serine availability for pyruvate synthesis (**Figure 23A**). However, providing serine in high quantities might be toxic to the cells.

Thus, to decrease serine oxidation for energy/reducing power synthesis, formate was added to the medium. Formate addition to VBS20 provides C1 units via the native Mis1p and Ade3 activity for intracellular C1 requirements and via its oxidation by the mtCbFdh1p module provides NADH for the intracellular requirement of reducing power and energy synthesis (**Figure 23B**). Thus, serine utilization for the C1 unit and energy synthesis can be reduced. Instead, serine would be predominantly available for biomass production via pyruvate synthesis (**Figure 23B**). Hence, in the media condition where serine and formate are provided, the VBS20 strain is leading comparatively to higher biomass production (**Figure 20B**).

Thus, these phenotypic observations can be interpreted as follows, when serine is supplied alone, the native Shm1/Shm2p degrades serine, impeding the pyruvate synthesis from serine.

Hence, a better way to establish the serine deamination module of the RGP without an energy module is to construct the S288c Δ*shm1*, Δ*shm2* genotype of S. cerevisiae and demonstrate faster or improved growth using serine alone as a carbon/energy source. Indeed even just an *SHM2* knockout should suffice to prevent the active serine oxidation in the cytoplasm as Shm2p contributes 95% of total *SHMTs* activity[100]. Independent of this strategy, the ability of VBS20 to grow on formate and serine is enough to move further for total biomass production from formate by constructing the serine and glycine synthesis modules of the RGP in the VBS20 strain.

6.7. Hypothetical effect of mutations on the RGP operation

The evolved VBS10 strains acquired several intergenic and intragenic mutations in the ALE process. Here, I discuss three important mutations found from the initial rounds of evolution (Ev01-Ev06) and maintained until the end, i.e., VBS10_Ev01-VBS10_Ev21. As explained in Section 5.4.1, Gdh1p::Δ36aa, Pet9p-Ile275Thr, and Idh1p-Pro299Gln genome modifications are considered to be important mutations in the coding regions of *GDH1*, *PET9*, and *IDH1*, encoding three proteins that are involved with the energy metabolism. These three mutations are expected to help the evolved VBS10 strains to improve the RGP operation directly or indirectly. The effect of Gdh1p::Δ36aa mutation on formate assimilation was confirmed via ^{13}C formate-based carbon tracing experiments, which showed that it leads to enhanced synthesis of CH_2-THF, an intermediate metabolite of the RGP.

The *GDH1::Δ108* bp mutation appeared very early, right from the Ev01. All the surviving cells in the Ev01 (that can produce serine and glycine from formate and CO_2) carry this mutation, and they retain it across the subsequent evolution stages. The function of the Gdh1p is to converts NH_3 and α-ketoglutarate into glutamate, consuming an NADPH for every reaction. In the tested culture condition, the medium's 2.5x high $(NH_4)_2SO_4$ concentration (100 mM) might enhance glutamate synthesis via Gdh1p, which may decrease intracellular NADPH levels. Upon the deletion of *GDH1::Δ108*, NADPH utilization for the glutamate synthesis is expected to decrease, as Gdh3p has less turnover, which is an isozyme for the alternative glutamate synthesis pathway

Labeling data of S288c_*Δgdh1::Δ108bp* showed increased formate assimilation compared to S288c_WT, primarily into methionine, serine, and alanine. This is consistent with the idea that the available NADPH is enhancing the CH_2-THF synthesis leading to the increased formate assimilation into the biomass. Additionally, as described in the review[130], the NADPH homeostasis mechanism based on metabolite signaling and regulations might improve the mitochondrial NADPH levels. This might enhance mitochondrial CH_2-THF synthesis, which might have influenced the reversal of the GCV system for partial glycine synthesis, especially

during the long incubation period of the ΔS strains. These ideas still need to be directly tested by measuring cellular NADPH and $NADP^+$ levels.

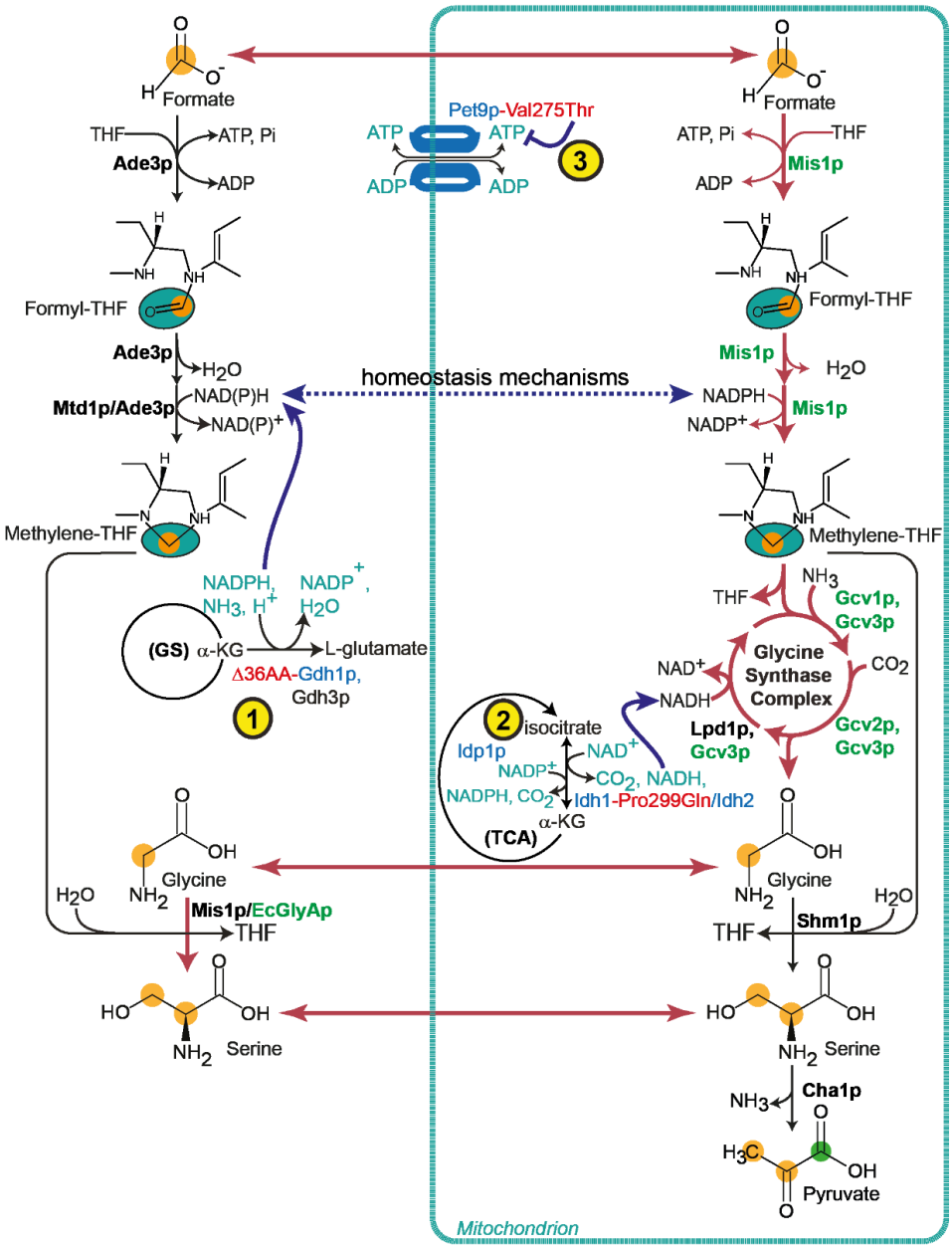


Figure 24: Hypothesized effect of mutations identified from NSG data in the RGP operation

Deletion of 108bp in the coding sequence of the *GDH1* gene is driving the flux towards CH_2-THF, leading to increase assimilation. The point mutation in the *IDH1* gene is speculated to produce more NADH which might help the regeneration of Gcv3p for the assimilation of the next round of glycine synthesis. This is speculated to improve the reversal of the GCV. The point mutation in Pet9p is speculated to reduce ATP transport into the cytoplasm in order to retain the ATP to operate formate ligation to the mitochondrial THF. This is speculated to improve the flux towards CHO-THF synthesis. **GS** – glyoxylate shunt; **α-KG** – α-ketoglutarate; **Green** color indicates the overexpression; **Black** color indicates the native expression of the enzyme. Mutated enzymes are indicated as blue-red colors; Blue- native enzyme carrying mutation and Red indicates the mutation; Purple indicates the possible effect of respective mutations.

In the S288c_gdh1::Δ108bp strain, ^{13}C labeled glycine is ~10% which is comparatively higher than S288c_WT, and it is subjected to high $(NH_4)_2SO_4$ in the medium as labeled glycine concentration decreases with a decrease in $(NH_4)_2SO_4$ concentration (**Figure S 13**). This indicates that the single labeled flux is more likely to be synthesized via the native GCV reversal activity. However, when high $(NH_4)_2SO_4$ is provided, serine deamination seems to be slightly inhibited, leading to less percentage of ^{13}C labeled alanine in the biomass (Figure S14). Nevertheless, whilst the exact mechanism is not fully established, the *GDH1::Δ108* mutation directly or indirectly leads to increased CH_2-THF synthesis that is the intermediate metabolite of the RGP.

PET9 is the major ATP/ADP translocator during aerobic growth and is located in the inner membrane of the mitochondria, where it catalyzes an electrogenic exchange of mitochondrial ATP^{4-} with ADP^{3-} from cytoplasm as this exchange leads to changes in extramitochondrial Mg(2+)[111]. Pet9p is essential, especially when yeast grows on non-fermentable carbon sources; several mutations were created and reported the effect on the growth of yeast while growing on glycerol[113]. Mutations at Arg252Ile, Arg253Ile, and Arg254Ile resulted in the loss of activity, but the Cys271Ser did not affect the function of Pet9p[113]. In this study, an Ile275Thr point mutation is identified in the Pet9p on the VBS10_Ev06 genome and retained across the evolution. I speculate that this point mutation might be regulating the ATP supply for the formate ligation to the THF system both in the mitochondria and cytoplasm, thus enhancing the growth rate of the evolved VBS10 strain by improving formate assimilation.

Another exciting point mutation that might influence the RGP operation is Idh1p-Pro299Gln. As explained in the Results section **5.4.1**., this mutation is present in the center of the AMP binding domain, Asn298-Pro299-Thr300. I speculate that mutation of the AMP binding domain to Asn298-Gln299-Thr300 might lead to conformational changes that modify activation of the catalytic Idh2p subunit, which catalyzes the interconversion of iso-citrate to α-ketoglutarate and, in the forward reaction, converts NAD^+ to NADH. Thus, I assume the Idh1p mutation might be helping in generating NADH, which may be supplied for the reduction of oxidized Gcv3p peptide upon release of glycine. The reduced Gcv3p is further involved in the cyclic assimilation of CH_2-THF and CO_2 for the synthesis of glycine.

As shown (**Figure 24**), the RGP requires 2 ATP, 2 NADPH, and 1 NADH molecule as energy and reducing power for synthesis of one serine molecule. Thus, the RGP is a reducing power- and energy-demanding pathway. As shown, *GDH1::Δ108* bp might be involved in the increased CH_2-THF synthesis, which on long-term incubation, might help yeast operate the RGP. The Pet9p- Ile275Thr might have improved the formate ligation to the mitochondrial and cytoplasmic THF, pushing the CH_2-THF synthesis for the assimilation into glycine and serine

synthesis, respectively, in the mitochondria and the cytoplasm. Finally, the Idh1p-Pro299Gln might be involved in the supply to NADH to reverse the GCV system, which might have improved the activity of the GSC. Together, these three mutations improved the overall formate-dependent growth of the evolved VBS10 strain between Ev01 to Ev06. Thus, I hypothesize that the initial mutation might have a role in activating the RGP. Later, two mutations might have enhanced the flux of the RGP by supporting energy and reducing power, leading to improved growth of the evolved VBS10 strains.

7. Conclusion

1. Constructed different ΔS strains in S288c and FL100 backgrounds of *S. cerevisiae.*
2. Established glycine to serine conversion modules in the ΔS strains of the S288c types that efficiently function in the RGP operable conditions.
3. Established glycine and serine synthesis from formate and CO_2 and improved it through ALE in the VBS10 and VBS16 strains.
4. Identified that GSC is operable at 2.5% CO_2 atmosphere in the VBS10_Ev21 strain, however, with a slow growth rate.
5. The best growing VBS10_Ev21 strain still requires high CO_2 for their optimum growth.
6. Identified mutations from genome sequencing of the evolved VBS10 strains.
7. Reverse engineered the *GDH1*::Δ108bp mutation in the unevolved VBS10 strain and reproduced the RGP-dependent growth in high CO_2 condition.
8. Validated that inactivation of *GDH1*::Δ108bp leads to improved formate assimilation.
9. Hypothesized the possible role of the mutations in operating the RGP.
10. Identified the native serine deaminase activity via the ^{13}C labeling experiments.
11. Demonstrated a complete RGP that converts formate to pyruvate via the ^{13}C formate labeling experiments.
12. The RGP has not yet been able to supply more than 10-17% of the pyruvate in the evolved VBS10 strains. The proportion of carbon needed for full growth that is coming from formate and CO_2 seems to be low.
13. Have not yet been able to establish a system for the generation of energy from formate in the evolved VBS10 strain, which means that glucose is still required in the medium. This, in turn, may hinder the establishment of experimental systems from selecting for the efficient conversion of serine to pyruvate in the evolved VBS10 strain. Thus, ALE in glucose limiting conditions failed to achieve full formatotrophy.
14. More generally, there may be a shortage of energy to support both formate assimilation and growth.
15. Engineered the serine deamination and energy production modules in the prototrophic VBS20 strain and established growth on serine and formate as the carbon and energy sources in this wild-type background without glucose in the medium.

8. Outlook

Even though all modules of the RGP, including the energy module, are successfully tested, growth of yeast on formate and CO_2 as the sole carbon and energy sources is still not accomplished. Nevertheless, full growth on formate and serine as the energy and carbon sources is established in the S288c_WT background strain, VBS20. Currently, the VBS20 strain carries *CHA1* and *mtCbFDH1* modules.

1. Further, to establish the pathway retrospectively, the glycine to serine conversion module that is tested, *EcSHMT*, can be targeted to mitochondria in the VBS20 strain. Then, establish growth on formate and glycine instead of serine in the medium without glucose.
2. Deleting *SHM2*, *GCV1*, or *GCV2* would be helpful to avoid degradation of glycine and serine supplied in the medium or synthesized intracellularly. If growth of yeast on glycine and formate is established, the further step would be to develop a glycine synthesis module.
3. In the next step, couple the overexpression module for glycine synthesis to the strain growing on glycine and formate and test for full formatotrophy.
4. Introduce *GDH1* mutation to push the RGP, overexpress *IDP1* for NADPH synthesis in mitochondrial, and evolve using formate/glycine or formate as the carbon and energy sources.
5. Test the effect of the *IDH1* and *PET9* point mutations on native formate assimilation pathways, and the RGP, if necessary, introduce these two mutations to speed up the evolution process.
6. Further, in an independent approach to improve the CO_2 assimilation via Gcv2p, directed evolution can be conducted by employing the evolved VBS10 strains under 2.5% CO_2 conditions using growth improvement as the readout.

9. References

1. Upadhyaya, S., Tiwari, S., Arora, N. & Singh, D. P. Microbial Protein: A Valuable Component for Future Food Security. *Microbes Environ. Manag.* 260–279 (2016) doi:10.13140/RG.2.1.1775.8801.

2. Fernando, S., Adhikari, S., Chandrapal, C. & Murali, N. Biorefineries: Current status, challenges, and future direction. *Energy and Fuels* **20**, 1727–1737 (2006).

3. Broecker, W. S. Climatic change: Are we on the brink of a pronounced global warming? *Science (80-.).* **189**, 460–463 (1975).

4. Cook, J. *et al.* Consensus on consensus: A synthesis of consensus estimates on human-caused global warming. *Environ. Res. Lett.* **11**, (2016).

5. Schneider, S. H. The global warming debate heats up: an analysis and perspective. *Bull. - Am. Meteorol. Soc.* **71**, 1292–1304 (1990).

6. The, O. N., Of, V., Climate, T. H. E., The, O. F. & Past, H. Quarterly journal. **XXVII**, (1901).

7. Vardy, M., Oppenheimer, M., Dubash, N. K., O'Reilly, J. & Jamieson, D. The Intergovernmental Panel on Climate Change: Challenges and Opportunities. *Annu. Rev. Environ. Resour.* **42**, 55–75 (2017).

8. Zaman, M. & Lee, J. H. Carbon capture from stationary power generation sources: A review of the current status of the technologies. *Korean J. Chem. Eng.* **30**, 1497–1526 (2013).

9. Odenberger, M. & Johnsson, F. Pathways for the European electricity supply system to 2050-The role of CCS to meet stringent CO2 reduction targets. *Int. J. Greenh. Gas Control* **4**, 327–340 (2010).

10. Biggs, S., Herzog, H., Reilly, J. & Jacoby, H. Economic Modeling of Co2 Capture and Sequestration. *Change* **13**, 16 (2000).

11. Monkman, S., Kline, J. & Cail, K. 15 th International Congress on the Chemistry of Cement Integrated capture and utilization of cement kiln CO2 to produce more sustainable concrete 15 th International Congress on the Chemistry of Cement. *15th Int. Congr. Chem. Cem.* (2019).

12. Monkman, S. & MacDonald, M. On carbon dioxide utilization as a means to improve the sustainability of ready-mixed concrete. *J. Clean. Prod.* **167**, 365–375 (2017).

13. Race, T. The RACE to. 24–26 (2011).

14. Kalyanarengan Ravi, N., Van Sint Annaland, M., Fransoo, J. C., Grievink, J. & Zondervan, E. Development and implementation of supply chain optimization framework for CO2 capture and storage in the Netherlands. *Comput. Chem. Eng.* **102**, 40–51 (2017).

15. Hepburn, C. *et al.* The technological and economic prospects for CO2 utilization and removal. *Nature* **575**, 87–97 (2019).

16. Huang, Z., Grim, R. G., Schaidle, J. A. & Tao, L. The economic outlook for converting CO2and electrons to molecules. *Energy Environ. Sci.* **14**, 3664–3678 (2021).

17. Whipple, D. T. & Kenis, P. J. A. Prospects of CO2 utilization via direct heterogeneous electrochemical reduction. *J. Phys. Chem. Lett.* **1**, 3451–3458 (2010).

18. Alsayegh, S. O. *et al.* Methanol Production Using Ultrahigh Concentrated Solar Cells: Hybrid Electrolysis and CO 2 Capture . *ACS Energy Lett.* **5**, 540–544 (2020).

19. Shafaat, H. S. & Yang, J. Y. Uniting biological and chemical strategies for selective CO2 reduction. *Nat. Catal.* **4**, 928–933 (2021).

20. Sullivan, I. *et al.* Coupling electrochemical CO2 conversion with CO2 capture. *Nat. Catal.* **4**, 952–958 (2021).

21. Yuan, M., Kummer, M. J. & Minteer, S. D. Strategies for Bioelectrochemical CO_2 Reduction. *Chem. - A Eur. J.* **25**, 14258–14266 (2019).

22. Masel, R. I. *et al.* An industrial perspective on catalysts for low-temperature CO_2 electrolysis. *Nat. Nanotechnol.* **16**, 118–128 (2021).

23. Tountas, A. A., Ozin, G. A. & Sain, M. M. Solar methanol energy storage. *Nat. Catal.* **4**, 934–942 (2021).

24. Whitaker, W. B. *et al.* Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. *Metab. Eng.* **39**, 49–59 (2017).

25. Kim, H. J. *et al.* Biological conversion of methane to methanol through genetic reassembly of native catalytic domains. *Nat. Catal.* **2**, 342–353 (2019).

26. Gregory, G. J., Bennett, R. K. & Papoutsakis, E. T. Recent advances toward the bioconversion of methane and methanol in synthetic methylotrophs. *Metab. Eng.* (2021) doi:10.1016/j.ymben.2021.09.005.

27. Kar, S., Goeppert, A. & Prakash, G. K. S. Integrated CO_2 Capture and Conversion to Formate and Methanol: Connecting Two Threads. *Acc. Chem. Res.* **52**, 2892–2903 (2019).

28. Vo, T. *et al.* Formate: An Energy Storage and Transport Bridge between Carbon Dioxide and a Formate Fuel Cell in a Single Device. *ChemSusChem* **8**, 3853–3858 (2015).

29. Crowther, T. W. *et al.* Mapping tree density at a global scale. *Nature* **525**, 201–205 (2015).

30. Bar-On, Y. M., Phillips, R. & Milo, R. The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6506–6511 (2018).

31. Gleizer, S. *et al.* Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. *Cell* **179**, 1255-1263.e12 (2019).

32. Gassler, T. *et al.* The industrial yeast Pichia pastoris is converted from a heterotroph into an autotroph capable of growth on CO_2. *Nat. Biotechnol.* (2019) doi:10.1038/s41587-019-0363-0.

33. Yishai, O., Lindner, S. N., Gonzalez de la Cruz, J., Tenenboim, H. & Bar-Even, A. The formate bio-economy. *Curr. Opin. Chem. Biol.* **35**, 1–9 (2016).

34. Weigmann, K. Fixing carbon. *EMBO Rep.* **20**, 2–5 (2019).

35. Cotton, C. A., Claassens, N. J., Benito-Vaquerizo, S. & Bar-Even, A. Renewable methanol and formate as microbial feedstocks. *Curr. Opin. Biotechnol.* **62**, 168–180 (2020).

36. Clomburg, J. M., Crumbley, A. M. & Gonzalez, R. Industrial biomanufacturing: The future of chemical production. *Science (80-.).* **355**, (2017).

37. Schrader, J. *et al.* Methanol-based industrial biotechnology: current status and future

perspectives of methylotrophic bacteria. *Trends Biotechnol.* **27**, 107–115 (2009).

38. Babel, W. The Auxiliary Substrate Concept: From simple considerations to heuristically valuable knowledge. *Eng. Life Sci.* **9**, 285–290 (2009).

39. Li, X., Yu, J., Jaroniec, M. & Chen, X. Cocatalysts for selective photoreduction of CO 2 into solar fuels. *Chem. Rev.* **119**, 3962–4179 (2019).

40. Choi, S. Y., Jeong, S. K., Kim, H. J., Baek, I. H. & Park, K. T. Electrochemical Reduction of Carbon Dioxide to Formate on Tin-Lead Alloys. *ACS Sustain. Chem. Eng.* **4**, 1311–1318 (2016).

41. Mandler, D. & Willner, I. Effective Photoreduction of CO2/HCO3to Formate Using Visible Light. *J. Am. Chem. Soc.* **109**, 7884–7885 (1987).

42. Moret, S., Dyson, P. J. & Laurenczy, G. Direct synthesis of formic acid from carbon dioxide by hydrogenation in acidic media. *Nat. Commun.* **5**, 1–7 (2014).

43. Jouny, M., Luc, W. & Jiao, F. General Techno-Economic Analysis of CO2 Electrolysis Systems. *Ind. Eng. Chem. Res.* **57**, 2165–2177 (2018).

44. Joó, F. Breakthroughs in hydrogen storage\-formic acid as a sustainable storage material for hydrogen. *ChemSusChem* **1**, 805–808 (2008).

45. Philips, M. F., Gruter, G. J. M., Koper, M. T. M. & Schouten, K. J. P. Optimizing the Electrochemical Reduction of CO2to Formate: A State-of-the-Art Analysis. *ACS Sustain. Chem. Eng.* **8**, 15430–15444 (2020).

46. Pletcher, D. The cathodic reduction of carbon dioxide - What can it realistically achieve? A mini review. *Electrochem. commun.* **61**, 97–101 (2015).

47. Li, H. *et al.* Integrated electromicrobial conversion of CO2 to higher alcohols. *Science (80-.).* **335**, 1596 (2012).

48. Rowaihi, I. S. Al *et al.* Poly(3-hydroxybutyrate) production in an integrated electromicrobial setup: Investigation under stress-inducing conditions. *PLoS One* **13**, 1–13 (2018).

49. Grunwald, S. *et al.* Kinetic and stoichiometric characterization of organoautotrophic growth of Ralstonia eutropha on formic acid in fed-batch and continuous cultures. *Microb. Biotechnol.* **8**, 155–163 (2015).

50. Song, Y. *et al.* Functional cooperation of the glycine synthasereductase and Wood-Ljungdahl pathways for autotrophic growth of Clostridium drakei. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 7516–7523 (2020).

51. Schiel-Bengelsdorf, B. & Dürre, P. Pathway engineering and synthetic biology using acetogens. *FEBS Lett.* **586**, 2191–2198 (2012).

52. Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci. U. S. A.* (2010) doi:10.1073/pnas.0907176107.

53. Bar-Even, A., Noor, E., Flamholz, A. & Milo, R. Design and analysis of metabolic pathways supporting formatotrophic growth for electricity-dependent cultivation of microbes. *Biochim. Biophys. Acta - Bioenerg.* **1827**, 1039–1047 (2013).

54. Kim, S. *et al.* Growth of E. coli on formate and methanol via the reductive glycine pathway. *Nat. Chem. Biol.* **16**, 538–545 (2020).

55. Claassens, N. J. *et al.* Replacing the Calvin cycle with the reductive glycine pathway

in Cupriavidus necator. *Metab. Eng.* **62**, 30–41 (2020).

56. Sánchez-Andrea, I. *et al.* The reductive glycine pathway allows autotrophic growth of Desulfovibrio desulfuricans. *Nat. Commun.* **11**, 1–12 (2020).

57. Figueroa, I. A. *et al.* Metagenomics-guided analysis of microbial chemolithoautotrophic phosphite oxidation yields evidence of a seventh natural CO2 fixation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E92–E101 (2018).

58. Bar-Even, A. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. *Biochemistry* **55**, 3851–3863 (2016).

59. Crowther, G. J., Kosály, G. & Lidstrom, M. E. Formate as the main branch point for methylotrophic metabolism in Methylobacterium extorquens AM1. *J. Bacteriol.* **190**, 5057–5062 (2008).

60. Tishkov, V. I. & Popov, V. O. Catalytic mechanism and application of formate dehydrogenase. *Biochem.* **69**, 1252–1267 (2004).

61. Claassens, N. J. *et al.* Engineering the Reductive Glycine Pathway: A Promising Synthetic Metabolism Approach for C1-Assimilation. (2022).

62. Mahadevan, R. & Henson, M. A. Genome-based modeling and design of metabolic interactions in microbial communities. *Comput. Struct. Biotechnol. J.* **3**, e201210008 (2012).

63. Parapouli, M., Vasileiadis, A., Afendra, A. S. & Hatziloukas, E. *Saccharomyces cerevisiae and its industrial applications*. *AIMS Microbiology* vol. 6 (2020).

64. Pereira, P. R., Freitas, C. S. & Paschoalin, V. M. F. Saccharomyces cerevisiae biomass as a source of next-generation food preservatives: Evaluating potential proteins as a source of antimicrobial peptides. *Compr. Rev. Food Sci. Food Saf.* **20**, 4450–4479 (2021).

65. Wang, Z., Zhuge, J., Fang, H. & Prior, B. A. Glycerol production by microbial fermentation: A review. *Biotechnol. Adv.* **19**, 201–223 (2001).

66. Hazelwood, L. A., Daran, J. M., Van Maris, A. J. A., Pronk, J. T. & Dickinson, J. R. The Ehrlich pathway for fusel alcohol production: A century of research on Saccharomyces cerevisiae metabolism. *Appl. Environ. Microbiol.* **74**, 2259–2266 (2008).

67. Nielsen, J., Larsson, C., van Maris, A. & Pronk, J. Metabolic engineering of yeast for production of fuels and chemicals. *Curr. Opin. Biotechnol.* **24**, 398–404 (2013).

68. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.* **31**, 335–341 (2013).

69. Shi, S. *et al.* Metabolic engineering of a synergistic pathway for n-butanol production in Saccharomyces cerevisiae. *Sci. Rep.* **6**, 1–10 (2016).

70. Nielsen, J. Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. *Bioengineered* **4**, 207–211 (2013).

71. Westfall, P. J. *et al.* Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 111–118 (2012).

72. Szczebara, F. M. *et al.* Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nat. Biotechnol.* **21**, 143–149 (2003).

73. Kjeldsen, T. Yeast secretory expression of insulin precursors. *Appl. Microbiol. Biotechnol.* **54**, 277–286 (2000).

74. Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science (80-.).* **349**, 1095–1100 (2015).

75. Gonzalez De La Cruz, J., Machens, F., Messerschmidt, K. & Bar-Even, A. Core Catalysis of the Reductive Glycine Pathway Demonstrated in Yeast. *ACS Synth. Biol.* **8**, 911–917 (2019).

76. Hermes, F. A. & Cronan, J. E. The role of the Saccharomyces cerevisiae lipoate protein ligase homologue, Lip3, in lipoic acid synthesis. *Yeast* (2013) doi:10.1002/yea.2979.

77. Jessop-Fabre, M. M. *et al.* EasyClone-MarkerFree: A vector toolkit for marker-less integration of genes into Saccharomyces cerevisiae via CRISPR-Cas9. *Biotechnol. J.* (2016) doi:10.1002/biot.201600147.

78. Hochrein, L. *et al.* AssemblX: A user-friendly toolkit for rapid and reliable multi-gene assemblies. *Nucleic Acids Res.* (2017) doi:10.1093/nar/gkx034.

79. Hurt, E. C., Pesold-Hurt, B., Suda, K., Oppliger, W. & Schatz, G. The first twelve amino acids (less than half of the pre-sequence) of an imported mitochondrial protein can direct mouse cytosolic dihydrofolate reductase into the yeast mitochondrial matrix. *EMBO J.* **4**, 2061–2068 (1985).

80. Gietz, R. D. Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol. Biol.* (2014) doi:10.1007/978-1-4939-0799-1_4.

81. Horwitz, A. A. *et al.* Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Syst.* **1**, 88–96 (2015).

82. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).

83. Ryan, O. W. *et al.* Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife* **3**, 1–15 (2014).

84. Mans, R. *et al.* CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Res.* **15**, 1–15 (2015).

85. Shaw, W. M. *et al.* Engineering a Model Cell for Rational Tuning of GPCR Signaling. *Cell* **177**, 782-796.e27 (2019).

86. Moqtaderi, Z. & Geisberg, J. V. Construction of mutant alleles in Saccharomyces cerevisiae without cloning: Overview and the Delitto perfetto method. *Curr. Protoc. Mol. Biol.* 1–17 (2013) doi:10.1002/0471142727.mb1310Cs104.

87. Sigma-Aldrich. Y1501 : without uracil Y1751 : without histidine Y1896 : without lysine Y1376 : without leucine Y1876 : without tryptophan Y0750 : without leucine and tryptophan Y1771 : without uracil , leucine , Y2146 : without histidine , leucine , Y2021 : without hist. 0–2 (2014).

88. You, L. *et al.* Metabolic pathway confirmation and discovery through 13c-labeling of proteinogenic amino acids. *J. Vis. Exp.* 1–6 (2012) doi:10.3791/3583.

89. Giavalisco, P. *et al.* Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* **68**, 364–376 (2011).

90. Yoshinaga, Y. & Dalin, E. RNase A Cleanup of DNA Samples. *Jt. Genome Inst.* 1–2 (2016).

91. Barrick, J. E. *et al.* Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* **15**, 1–17 (2014).

92. Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science (80-.).* **355**, 1040–1044 (2017).

93. Pretorius, I. S. & Boeke, J. D. Yeast 2.0-connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Res.* **18**, 1–15 (2018).

94. Richardson, S. M. *et al. Design of a synthetic yeast genome Downloaded from*. *Science* vol. 355 http://science.sciencemag.org/ (2017).

95. Albers, E., Laizé, V., Blomberg, A., Hohmann, S. & Gustafsson, L. Ser3p (Yer081wp) and Ser33p (Yil074cp) are phosphoglycerate dehydrogenases in Saccharomyces cerevisiae. *J. Biol. Chem.* **278**, 10264–10272 (2003).

96. Ulane, R. & Ogur, M. Genetic and physiological control of serine and glycine biosynthesis in Saccharomyces. *J. Bacteriol.* **109**, 34–43 (1972).

97. Takada, Y. & Noguchi, T. Characteristics of alanine:glyoxylate aminotransferase from Saccharomyces cerevisiae, a regulatory enzyme in the glyoxylate pathway of glycine and serine biosynthesis from tricarboxylic acid-cycle intermediates. *Biochem. J.* **231**, 157–163 (1985).

98. McNeil, J. B. *et al.* Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. *J. Biol. Chem.* **269**, 9155–9165 (1994).

99. Schlösser, T., Gätgens, C., Weber, U. & Stahmann, K. P. Alanine: Glyoxylate aminotransferase of Saccharomyces cerevisiae-encoding gene AGX1 and metabolic significance. *Yeast* **21**, 63–73 (2004).

100. Kastanos, E. K., Woldman, Y. Y. & Appling, D. R. Role of mitochondrial and cytoplasmic serine hydroxymethyltransferase isozymes in de Novo purine synthesis in Saccharomyces cerevisiae. *Biochemistry* **36**, 14956–14964 (1997).

101. Schï osser, T., Gätgens, C., Weber, U. & Stahmann, K.-P. Alanine : glyoxylate aminotransferase of Saccharomyces cerevisiae-encoding gene AGX1 and metabolic significance. *Yeast* **21**, 63–73 (2004).

102. Subramanian, M. *et al.* Transcriptional regulation of the one-carbon metabolism regulon in Saccharomyces cerevisiae by Bas1p. *Molecular Microbiology* vol. 57 53–69 (2005).

103. Sinclair, D. A. & Dawes, I. W. Genetics of the synthesis of serine from glycine and the utilization of glycine as sole nitrogen source by Saccharomyces cerevisiae. *Genetics* **140**, 1213–1222 (1995).

104. Pasternack, L. B., Laude, D. A. & Appling, D. R. 13C NMR Detection of Folate-Mediated Serine and Glycine Synthesis in Vivo in Saccharomyces Cerevisiaet. *Biochemistry* (1992) doi:10.1021/bi00152a005.

105. Piper, M. D. W., Hong, S. P., Eißing, T., Sealey, P. & Dawes, I. W. Regulation of the yeast glycine cleavage genes is responsive to the availability of multiple nutrients. *FEMS Yeast Res.* **2**, 59–71 (2002).

106. Piper, M. D., Hong, S. P., Ball, G. E. & Dawes, I. W. Regulation of the balance of one-

carbon metabolism in Saccharomyces cerevisiae. *J. Biol. Chem.* **275**, 30987–30995 (2000).

107. Monerawela, C., James, T. C., Wolfe, K. H. & Bond, U. Loss of lager specific genes and subtelomeric regions define two different Saccharomyces cerevisiae lineages for Saccharomyces pastorianus Group I and II strains. *FEMS Yeast Res.* **15**, 1–11 (2015).

108. Stewart, G. G., Hill, A. E. & Russell, I. 125th Anniversary Review: Developments in brewing and distilling yeast strains. *J. Inst. Brew.* **119**, 202–220 (2013).

109. DeLuna, A., Avendaño, A., Riego, L. & González, A. NADP-glutamate dehydrogenase isoenzymes of Saccharomyces cerevisiae: Purification, kinetic properties, and physiological roles. *J. Biol. Chem.* **276**, 43775–43783 (2001).

110. Mara, P., Fragiadakis, G. S., Gkountromichos, F. & Alexandraki, D. The pleiotropic effects of the glutamate dehydrogenase (GDH) pathway in Saccharomyces cerevisiae. *Microb. Cell Fact.* **17**, 1–13 (2018).

111. Adrian, G. S., McCammon, M. T., Montgomery, D. L. & Douglas, M. G. Sequences required for delivery and localization of the ADP/ATP translocator to the mitochondrial inner membrane. *Mol. Cell. Biol.* **6**, 626–634 (1986).

112. Traba, J., Froschauer, E. M., Wiesenberger, G., Satrústegui, J. & Del Arco, A. Yeast mitochondria import ATP through the calcium-dependent ATP-Mg/Pi carrier Sal1p, and are ATP consumers during aerobic growth in glucose. *Mol. Microbiol.* **69**, 570–585 (2008).

113. JMB-1993-Nelson-230-1159.pdf.

114. Cupp, J. R. & McAlister-Henn, L. Kinetic Analysis of NAD+-Isocitrate Dehydrogenase with Altered Isocitrate Binding Sites: Contribution of IDH1 and IDH2 Subunits to Regulation and Catalysis. *Biochemistry* **32**, 9323–9328 (1993).

115. Panisko, E. A. & McAlister-Henn, L. Subunit interactions of yeast NAD+-specific isocitratec dehydrogenase. *J. Biol. Chem.* **276**, 1204–1210 (2001).

116. Gachotte, D., Barbuch, R., Gaylor, J., Nickel, E. & Bard, M. Characterization of the Saccharomyces cerevisiae ERG26 gene encoding the C-3 sterol dehydrogenase (C-4 decarboxylase) involved in sterol biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 13794–13799 (1998).

117. Wang, D., Lai, F. L. & Gao, F. Ori-Finder 3: A web server for genome-wide prediction of replication origins in Saccharomyces cerevisiae. *Brief. Bioinform.* **22**, 1–13 (2021).

118. Ikuchi, B. G. K., Otokawa, Y. M., Ã, Ã. T. Y. O. & Ã, K. H. I. Glycine cleavage gcvTHP review. **84**, 246–263 (2008).

119. Subramanian, M. *et al.* Transcriptional regulation of the one-carbon metabolism regulon in Saccharomyces cerevisiae by Bas1p. *Mol. Microbiol.* **57**, 53–69 (2005).

120. Yang, Y. & Meier, U. T. Genetic interaction between a chaperone of small nucleolar ribonucleoprotein particles and cytosolic serine hydroxymethyltransferase. *J. Biol. Chem.* **278**, 23553–23560 (2003).

121. Zhong, Z. *et al.* Automated Continuous Evolution of Proteins in Vivo . *ACS Synth. Biol.* (2020) doi:10.1021/acssynbio.0c00135.

122. Tan, Z. L. *et al.* In vivo continuous evolution of metabolic pathways for chemical production. *Microb. Cell Fact.* **18**, 1–19 (2019).

123. Petersen, J. G. L., Kielland-brandt, M. C., Nilsson-tillgren, T., Bornaes, C. & Holmberg, S. Molecular Genetics. **534**, 527–534 (1988).

124. Bornaes, C., Ignjatovic, M. W., Schjerling, P., Kielland-Brandt, M. C. & Holmberg, S. A regulatory element in the CHA1 promoter which confers inducibility by serine and threonine on Saccharomyces cerevisiae genes. *Mol. Cell. Biol.* **13**, 7604–7611 (1993).

125. Holmberg, S. & S. Activates Transcription Via Serine/Threonine Response Elements. *Genetics* **144**, 467–478 (1996).

126. Overkamp, K. M. *et al.* Functional analysis of structural genes for NAD+ -dependent formate dehydrogenase in Saccharomyces cerevisiae. *Yeast* vol. 19 509–520 (2002).

127. Hou, J., Scalcinati, G., Oldiges, M. & Vemuri, G. N. Metabolic impact of increased NADH availability in saccharomyces cerevisiae. *Appl. Environ. Microbiol.* **76**, 851–859 (2010).

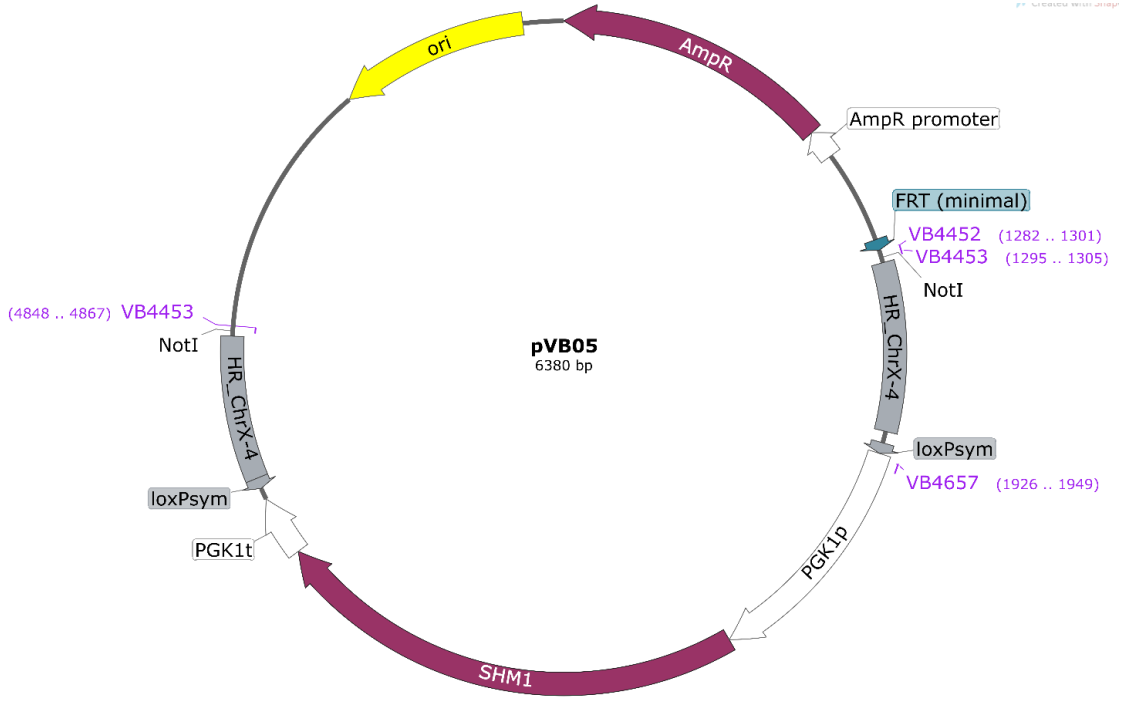
128. Sakai, Y., Murdanoto, A. P., Konishi, T., Iwamatsu, A. & Kato, N. Regulation of the formate dehydrogenase gene, FDH1, in the methylotrophic yeast Candida boidinii and growth characteristics of an FDH1- disrupted strain on methanol, methylamine, and choline. *J. Bacteriol.* **179**, 4480–4485 (1997).

129. Yurimoto, H., Oku, M. & Sakai, Y. Yeast methylotrophy: Metabolism, gene regulation and peroxisome homeostasis. *Int. J. Microbiol.* **2011**, (2011).

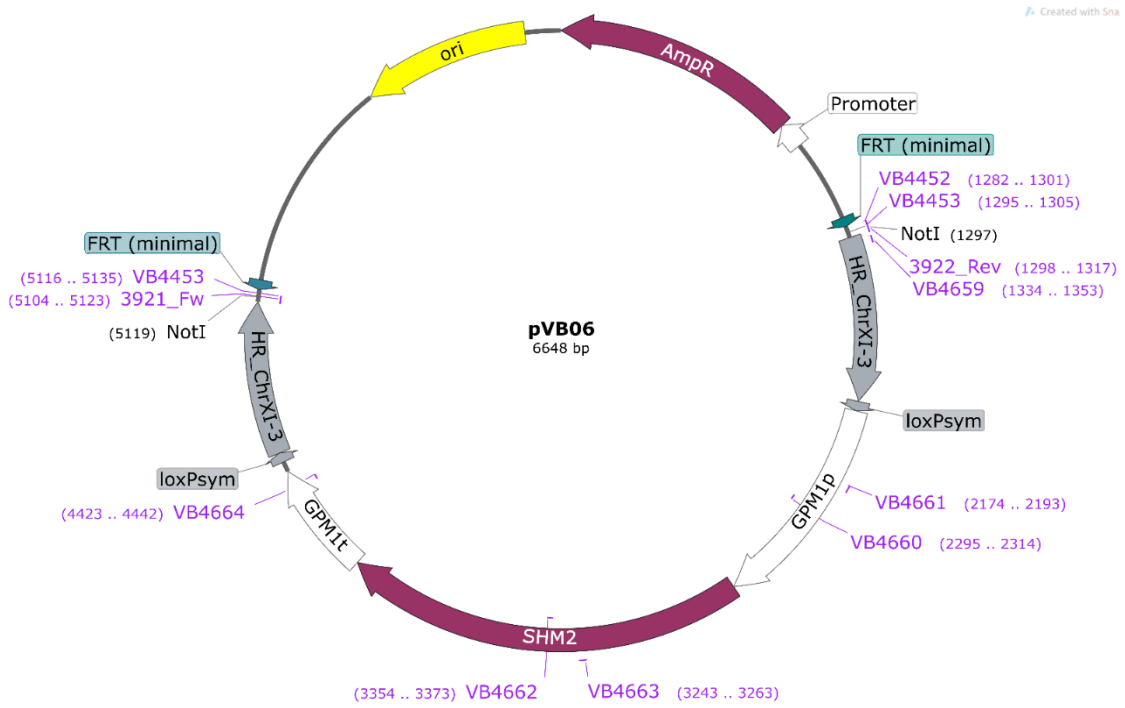
130. Xiao, W., Wang, R. S., Handy, D. E. & Loscalzo, J. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxidants Redox Signal.* **28**, 251–272 (2018).

10. Supplementary information

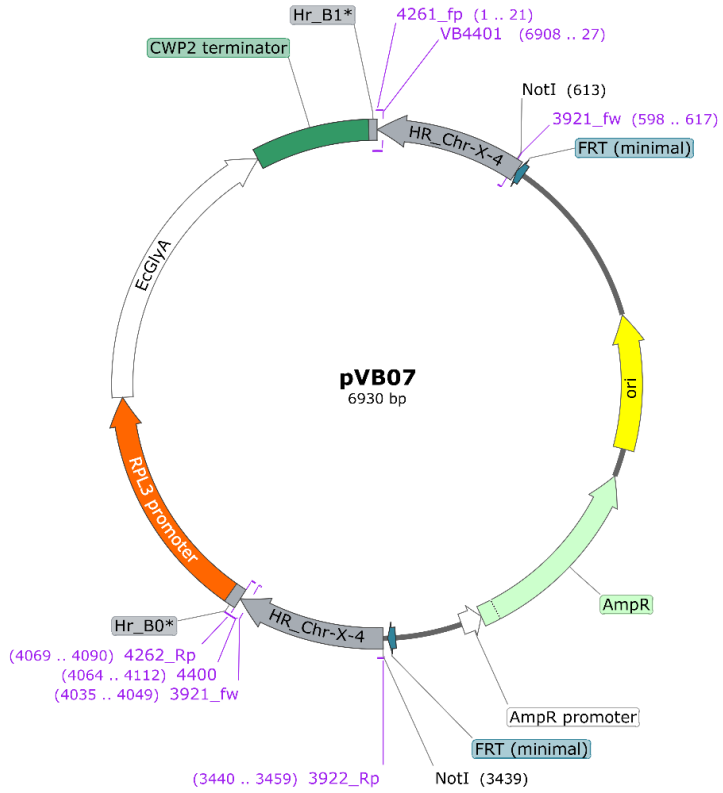
10.1. Supplemental plasmid maps S1-S8



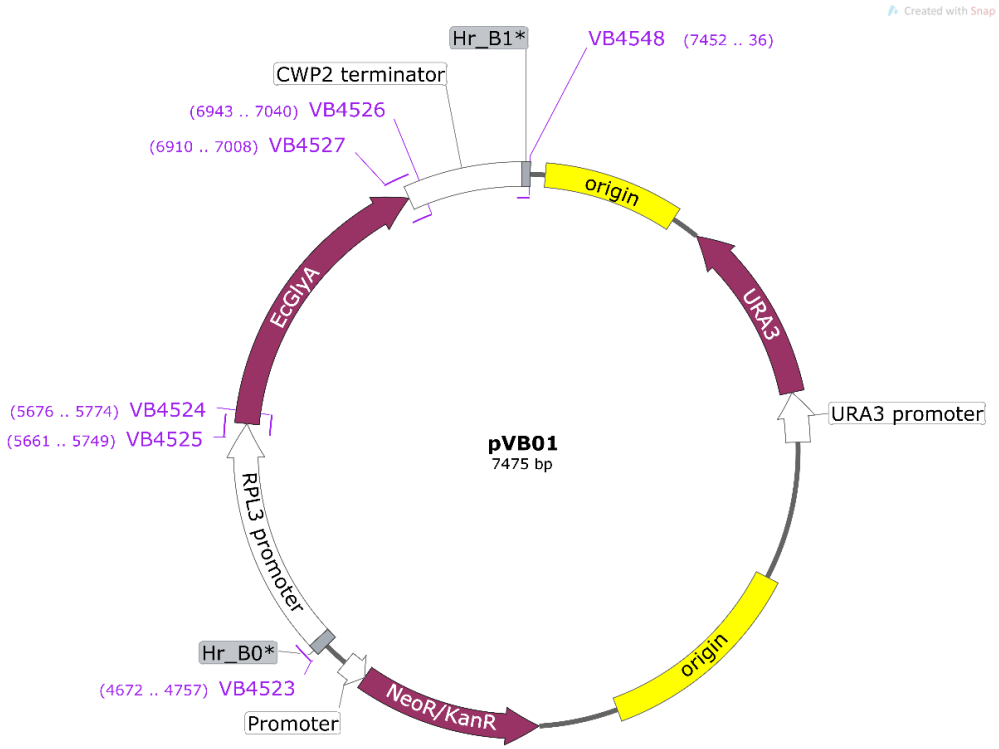
Map S 1: Map of pVB05 plasmid
Map of pVB05 plasmid map with *SHM1* expression module under the *PGK1* promoter and terminator cloned between ChrX-4 integration sites.



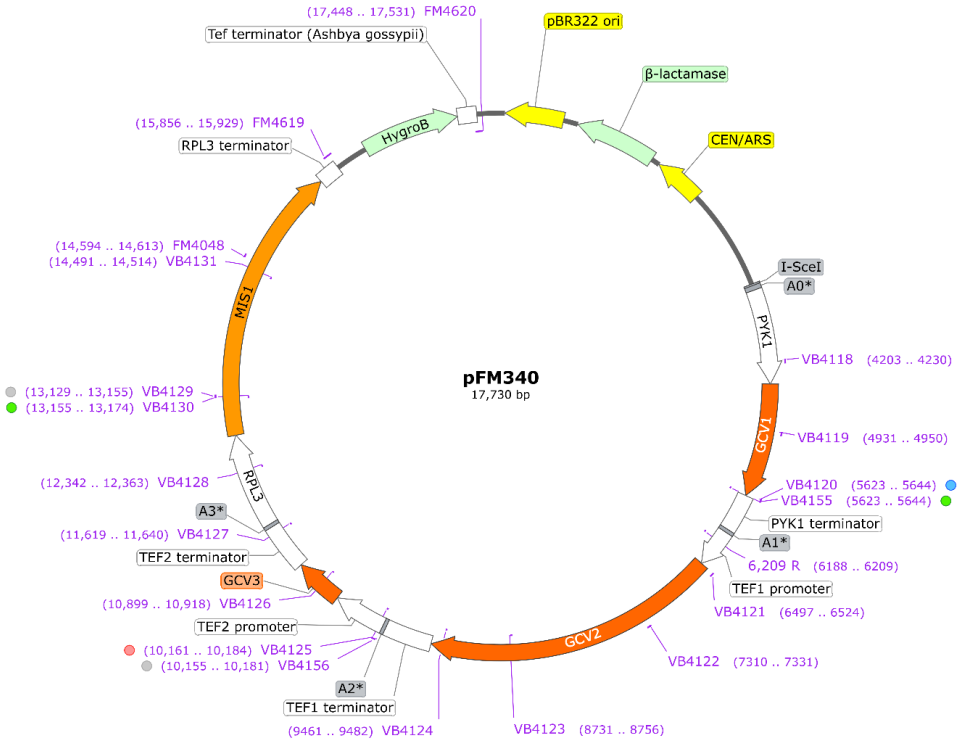
Map S 2: Map of pVB05 plasmid
Map of pVB06 plasmid with *SHM2* expression module under the *GPM1* promoter and terminator cloned between ChrXI-3 integration sites.



Map S 3: Map of pVB07 plasmid
Map of pVB07 plasmid with *EcGlyA* expression module under the *RPL3* promoter and *CWP2* terminator cloned between ChrX-4 integration sites.

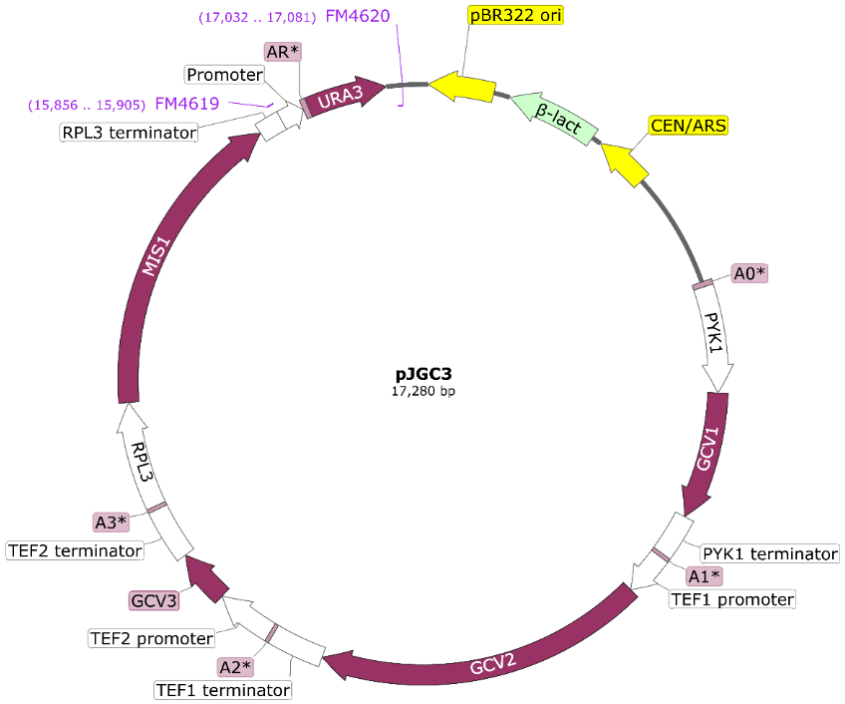


Map S 4: Map of pVB01 plasmid
Plasmid map of pVB01 with *EcGlyA* expression module under the *RPL3* promoter and *CWP2* terminator.



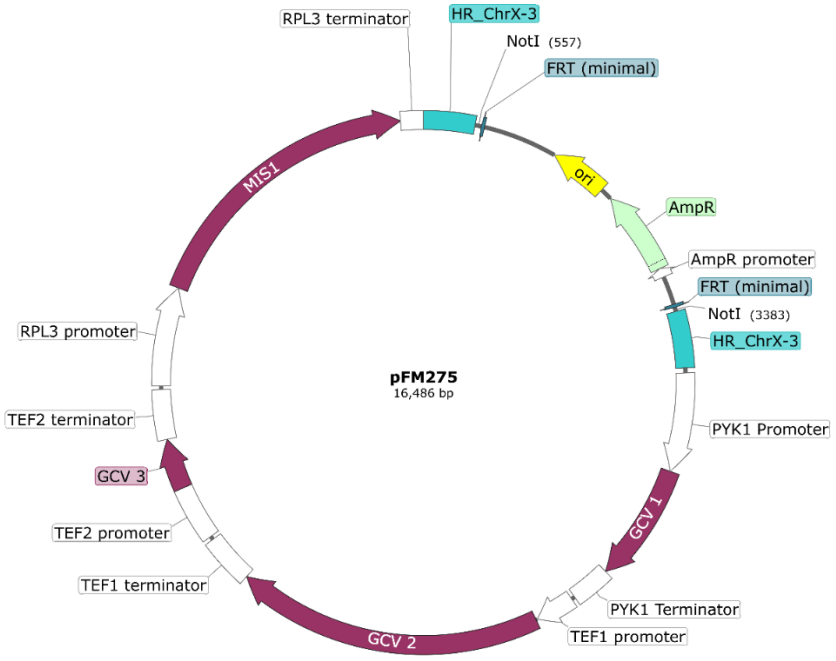
Map S 5: Plasmid map of pFM340

Map of pFM340 plasmid with *GCV1*, *GCV2*, *GCV3*, and *MIS1* genes under the promoter and terminators of *PYK1*, *TEF1*, *TEF2*, and *RPL3,* respectively. pFM340 carries a Hygromycin resistance cassette for selection in the yeast system. It is a centromere plasmid with CEN/ARS Ori.



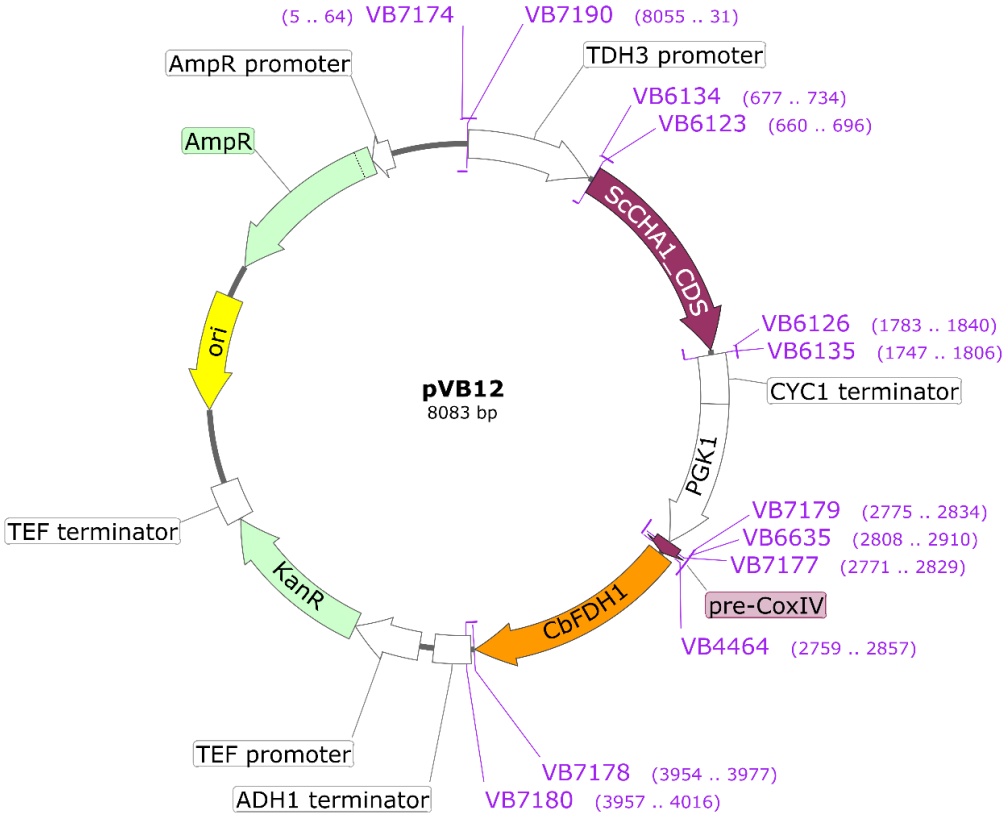
Map S 6: Plasmid map of pJGC3

Map of pJGC3 plasmid with expression cassettes of *GCV1*, *GCV2*, *GCV3*, and *MIS1* under the promoter and terminators of *PYK1*, *TEF1*, *TEF2*, and *RPL3,* respectively. pFM340 carries a URA3 selection marker to maintain in the yeast system. It is a centromere plasmid with CEN/ARS Ori.



Map S 7: Plasmid map of pFM275

Map of pFM275 plasmid with expression cassettes of *GCV1*, *GCV2*, *GCV3*, and *MIS1* under the promoter and terminators of *PYK1*, *TEF1*, *TEF2*, and *RPL3* cloned between ChrX-3 locus genome integration sites.



Map S 8: Plasmid map of pVB13

Map of pVB12 plasmid with expression cassettes of *CHA1* under *TDH3* promoter and *CYC1* terminator and the expression cassette of Pre-CoxIV (mitochondrial tag) fused *CbFDH1* (Cb-*Candida boindii*) gene under the *PGK1* promoter and *ADH1* terminator.

10.2. Supplemental figures S1-S12

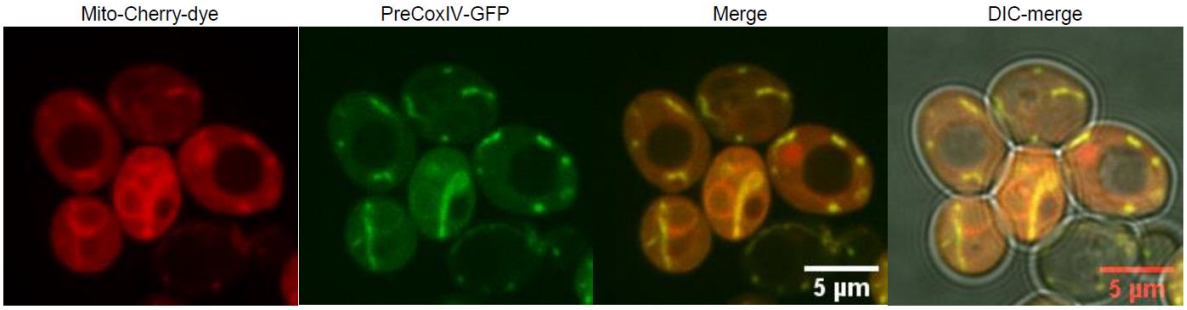


Figure S 1. Fluorescent microscopic images of yeast cells showing *Pre-CoxIV* fused *GFP* localized in mitochondria Panel-1 –Mitochondria of yeast cells stained with Mito-cherry dye under red fluorescent channel; **Panel-2** – *PreCoxIV* fused *GFP* expression in yeast targeted to mitochondria under the green fluorescent channel. **Panel-3** – Overlay of *GFP* expression with Mito-cherry staining of mitochondrial. The yellow color indicates colocalization of the *GFP* to mitochondria. Panel-4 – merge of the DIC image of yeast cells with red and green fluorescent channels.

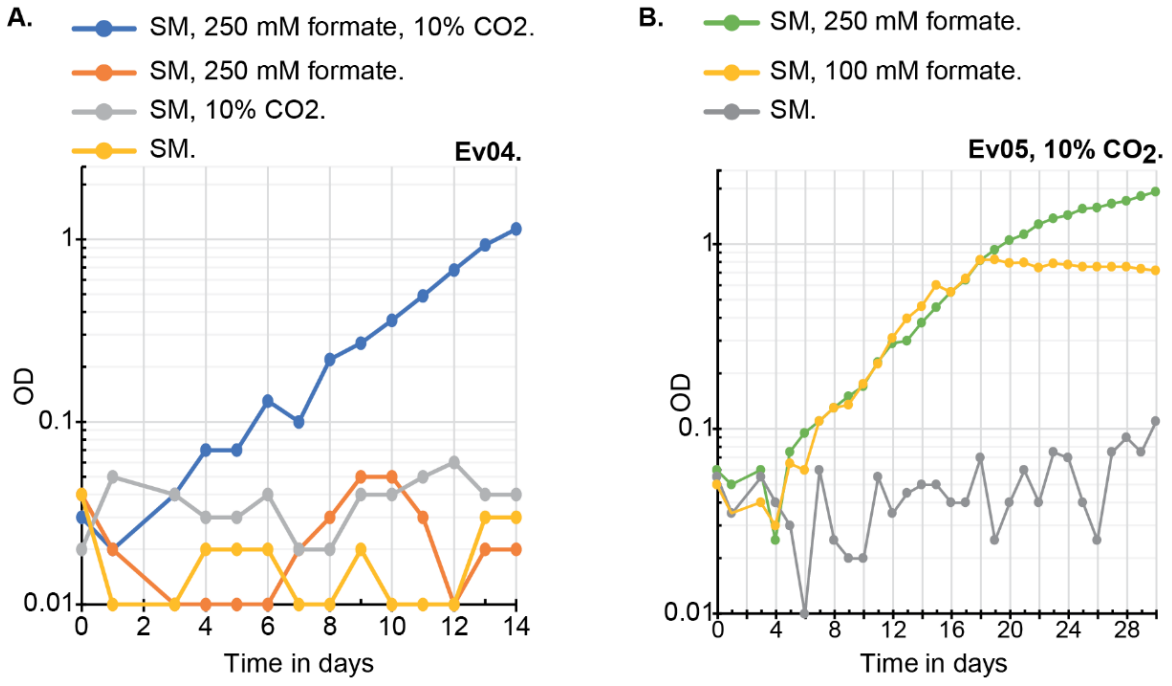


Figure S 2: Growth curves of VBS10 strain from Ev04 and Ev05: A. Growth of VBS10 strain from Ev04 in 10% high CO2 and ambient CO2 conditions in SM medium with and without formate supplementation**. B.** Growth curves of VBS10 strain in Ev05 in SM medium with 100 mM and 250 mM formate. 100 mM formate condition led to lower OD_{600} compared to 250 mM formate condition, but no growth was observed without formate. **SM**= 1 YNB, 100 mM of glucose and 100 mM of $(NH_4)_2SO_4$.

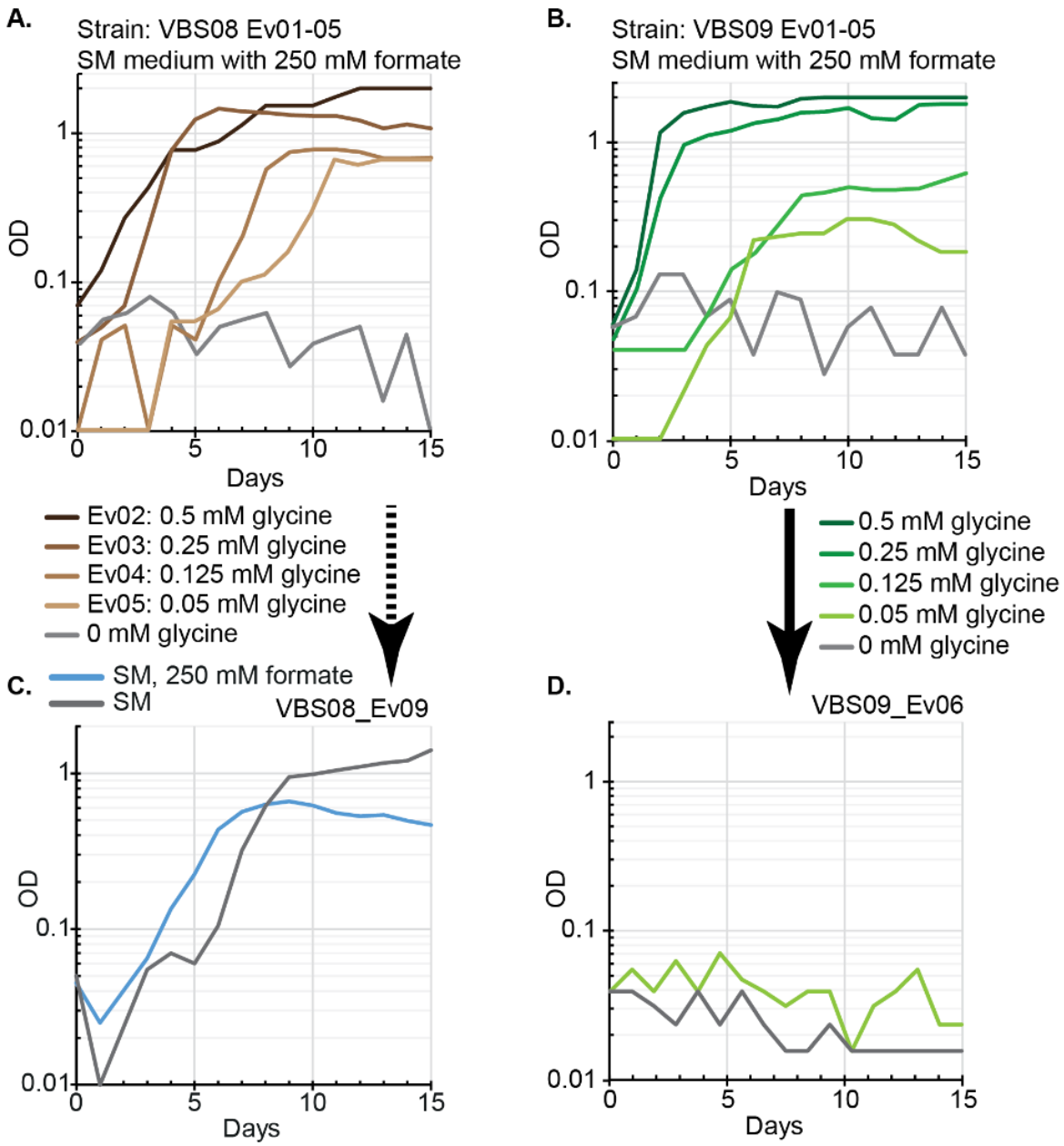


Figure S 3: Evolution of VBS08 and VBS09 in glycine limiting process

A. Growth curves of VBS08 strain in lower glycine concentrations from 1 mM to 0.05 mM concentration in SM medium with and without formate from Ev01-Ev05. **B.** Growth curves of VBS09 strain in lower glycine concentrations from 1 mM to 0.05 mM concentration in SM medium with and without formate. **C.** Growth curves of VBS08 strain in SM medium with and without formate. **D.** Growth curves of VBS09 strain in SM medium supplemented with formate in the presence or absence of 0.05 mM glycine. **SM**= 1x YNB, 100 mM ammonium sulfate, 100 mM glucose, formate and glycine concentrations as indicated in the graphs.

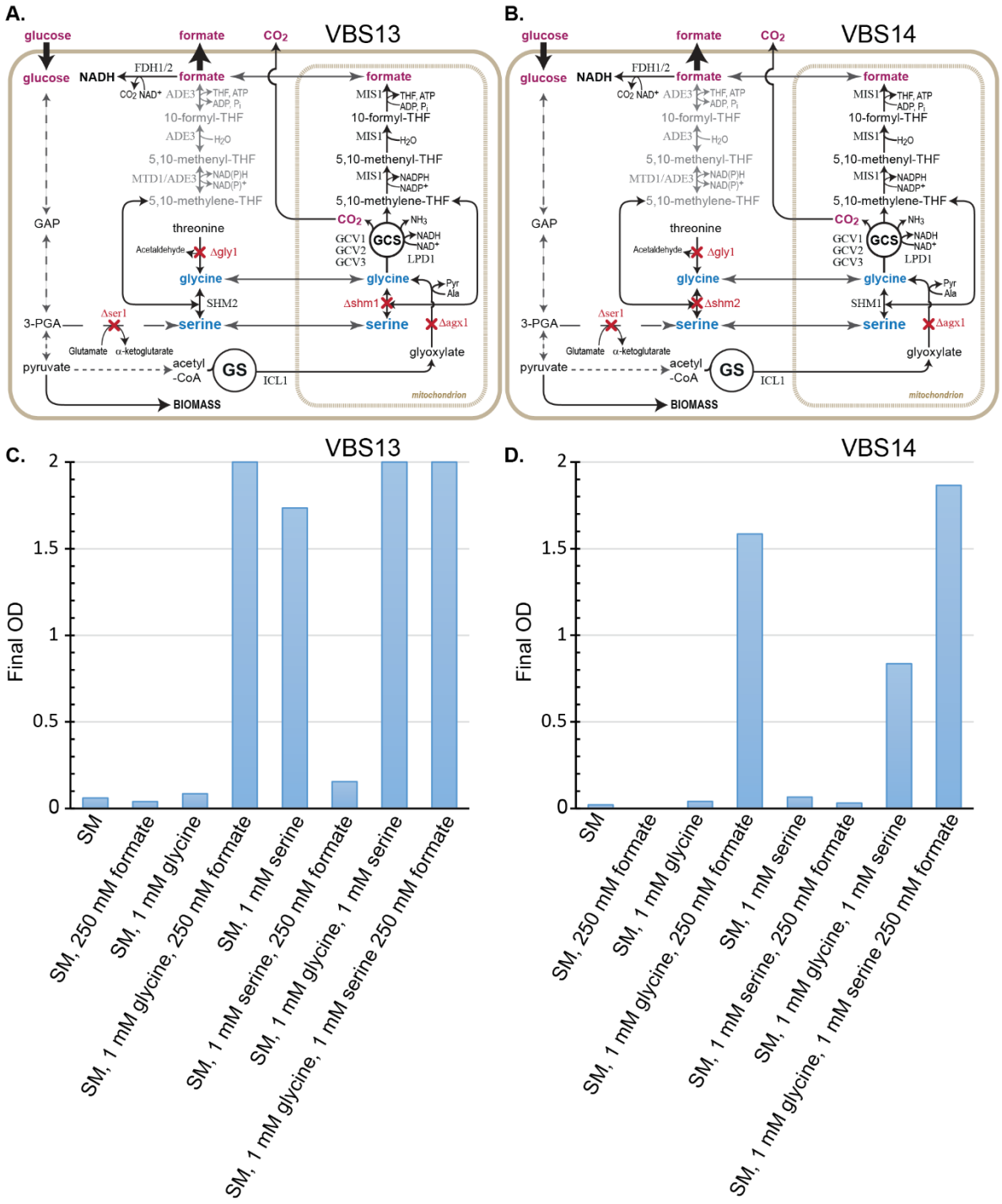


Figure S 4: Serine biosensor strains and metabolic phenotypes of FL100 strain

A. Metabolic scheme of the VBS13 serine biosensor strain. **B.** Metabolic scheme of the VBS14 serine biosensor strain. **C.** Final ODs of VBS13 strain in the presence of glycine, serine, and formate combinations in the SM medium provided with glucose as the primary carbon source. **D.** Final OD_{600} of VBS13/14 strains in the presence of glycine, serine, and formate combinations in the SM medium provided with glucose as the primary carbon source. **GS**- glyoxylate shunt; **SM**= 1x YNB, 100 mM glucose, 100 mM $(NH_4)_2SO_4$, formate and glycine as indicated in the graphs.

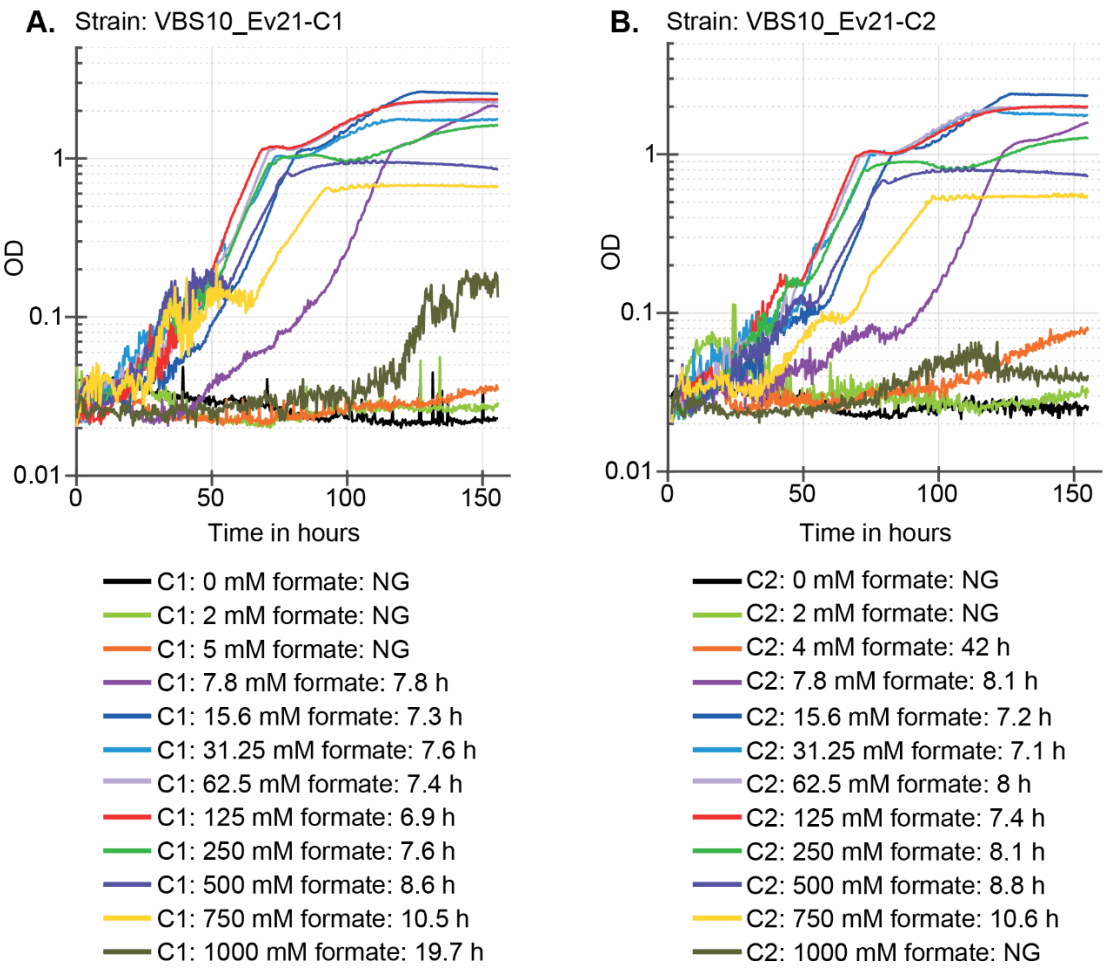


Figure S 5: Growth curves of VBS10_Ev21 strains in the formate gradient
Growth curves of two different clones from VBS10_Ev21 with various concentrations of formate ranging from 2 mM to 1000 mM. Medium: **SM**= 1x YNB, 100 mM $(NH_4)_2SO_4$, 50 mM glucose;

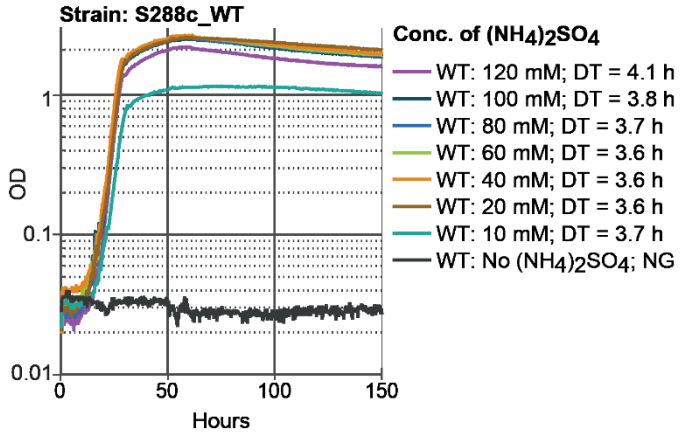


Figure S 6: Growth of S288c_WT strain in ammonium sulfate gradient
Growth curves of S288c_WT with $(NH_4)_2SO_4$ concentrations from 10 mM to 120 mM in SM medium. **SM**= 1x YNB, 50 mM glucose, 100 mM formate.

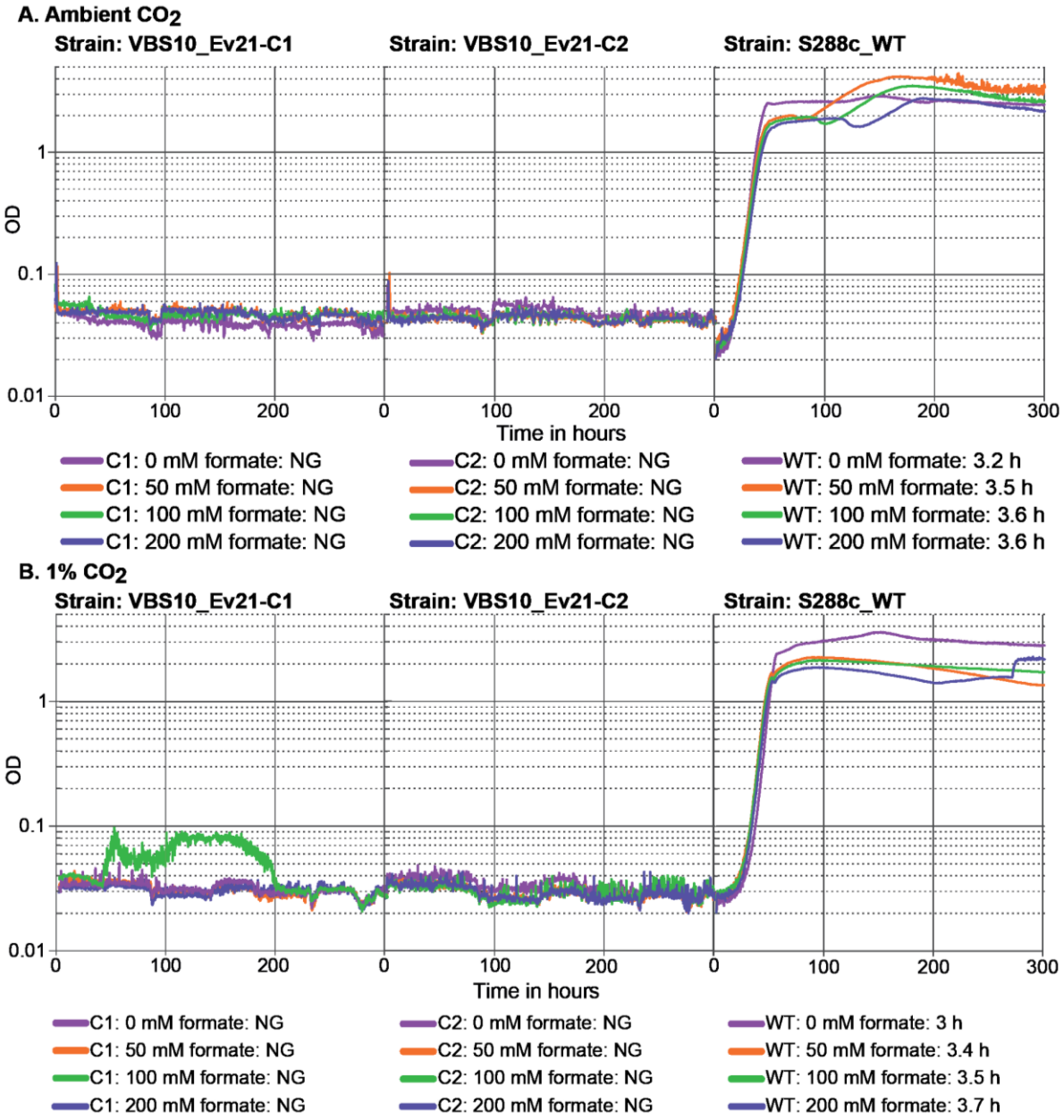


Figure S 7: Growth curves of VBS10_Ev21 and S288c_WT in ambient

A. Growth of the VBS10_Ev21 strains and S288c_WT in different concentrations of formate in ambient CO_2 conditions. **B.** Growth of the VBS10_Ev21 strains and S288c_WT in the presence of a different concentration of formate in a 1% CO_2 atmosphere. **SM** medium with 1x YNB, 100 mM $(NH_4)_2SO_4$ and 50 mM glucose.

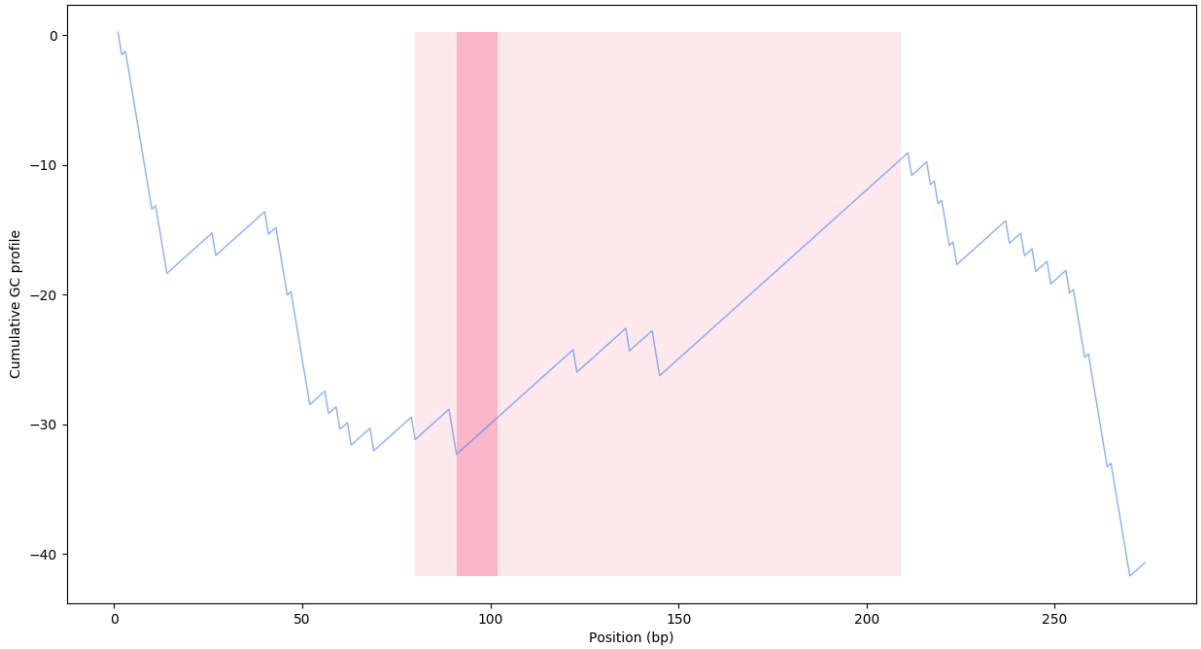


Figure S 8: Prediction of Ori with point mutation

A total 250 bp in the upstream and downstream area of the point mutation were scanned and highlighted the presence of Autonomously replicating sequence (ARS) in the mitochondrial chromosome. Stronger color indicates the location of Ori in between 90-100 bp.

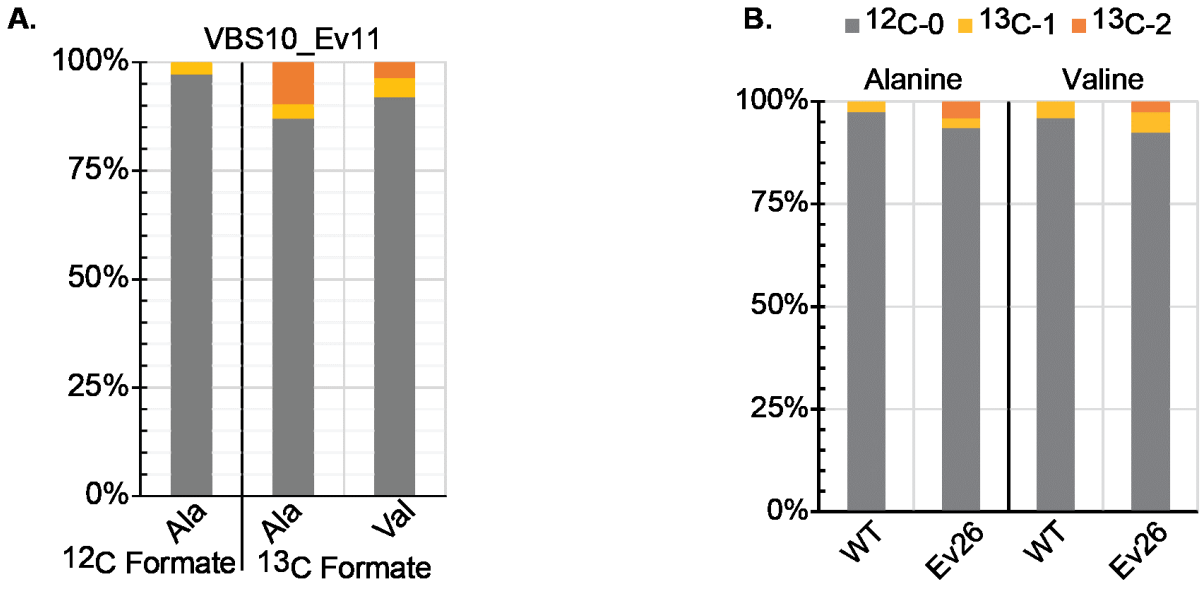


Figure S 9: First evidence of pyruvate biosynthesis from formate

Biomass of the VBS10_Ev11 fed with ^{13}C formate and ^{12}C glucose was collected, washed with sterile water and acid hydrolyzed at 95ºC to amino acids. Resuspended the hydrolyzed biomass in sterile water, filtered and sent for LC-MS analysis of amino acids. Data analysed for labeling pattern in the alanine and valine. **A.** Labeling data shows 10% of $^{13}C_2$-alanine and ~4% of $^{13}C_2$-valine from VBS_Ev11 biomass incubated fed with ^{13}C-formate. **B.** Comparison of labeling data between WT and VBS10_Ev26 strain, ~4% of alanine shows the $^{13}C_2$ carbon and ~2.5% of valine carrying $^{13}C_2$ carbons.

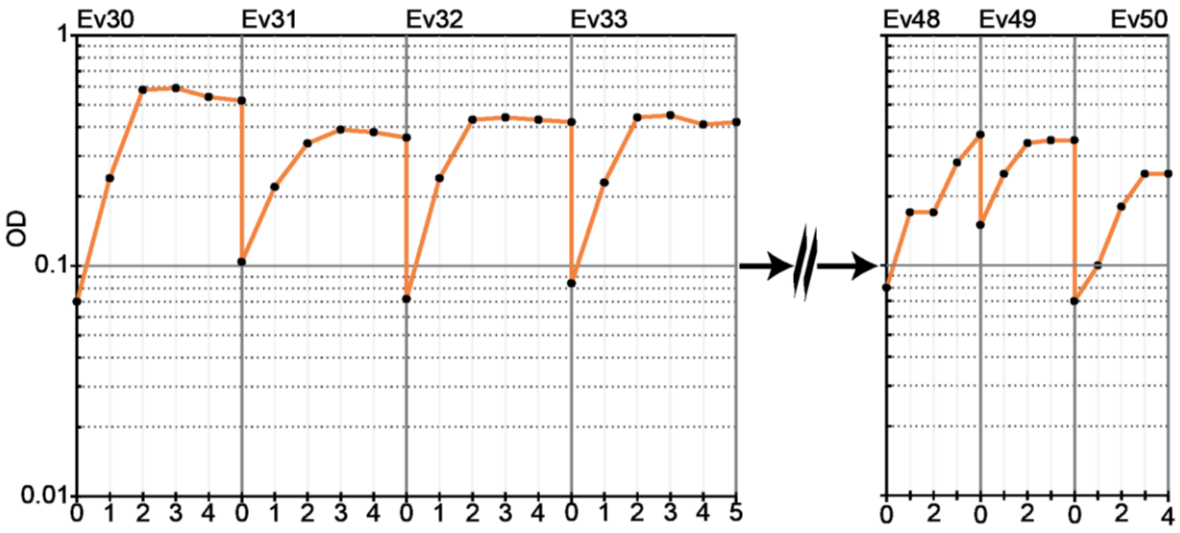


Figure S 10: Evolution of the VBS10_Ev29 strain in glucose limiting condition

Evolution of VBS10_Ev29 strains in limited glucose condition in SM medium supplemented with 250mM formate and 2mM glucose from Ev30 to Ev50. Strain did not evolve as growth was not improved.

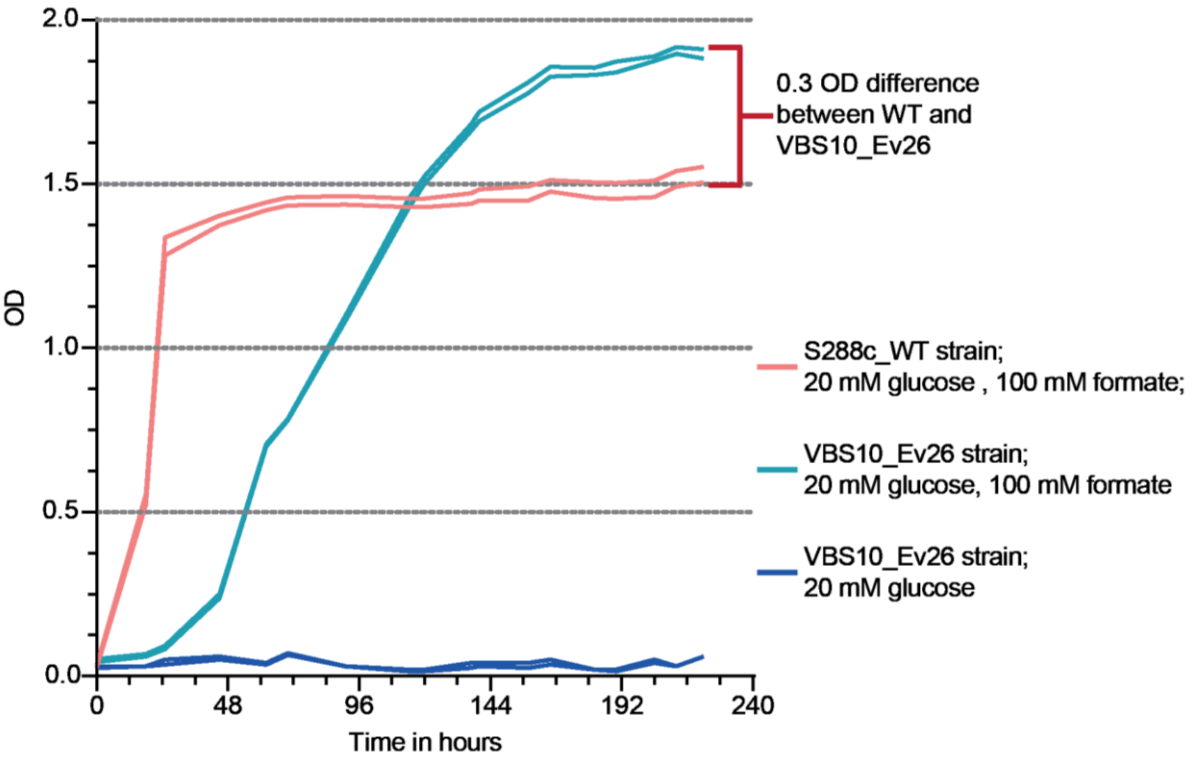


Figure S 11: Growth comparison of VBS10_Ev26 with S288c_WT

Growth curves of the VBS10_Ev26 and S288c WT strain in SM medium supplemented with 100 mM formate and 20 mM glucose. Increased biomass by 0.3 OD is observed in the VBS10_Ev26 strain compared to the WT when grown on a limited carbon source.

Figure S 12: Pairwise alignment of the *FDH2*/*FDH1* coding sequences

Coding sequence of FDH2/FDH1 of S288c strain were aligned using Blast-Global align tool from NCBI. The *FDH2* gene lost ~360 bp of coding sequence gene along with two miss-match in the aligned sequences compared *FDH1*.

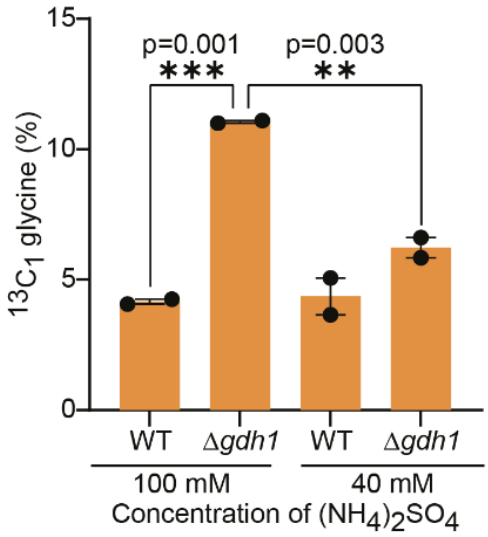


Figure S 13: Influence of $(NH_4)_2SO_4$ on glycine synthesis in gdh1 mutant

Even though the overall percentage of glycine carrying single ^{13}C carbon is low, a variation is observed with respect to $(NH_4)_2SO_4$ concentration in the medium between the gdh1 mutant and WT control. Significantly higher single-labeled glycine is observed when 100 mM $(NH_4)_2SO_4$ is provided in the medium compared to 40 mM $(NH_4)_2SO_4$.

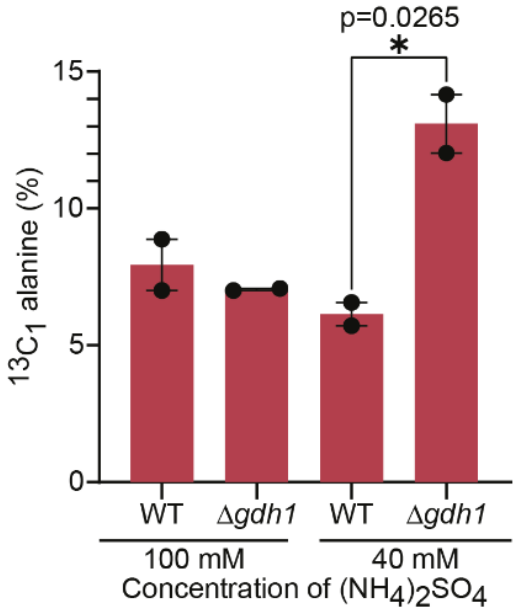


Figure S 14: Influence of (NH4)2SO4 on the alanine synthesis via Cha1p

Even though the overall percentage of alanine carrying single ^{13}C carbon is low, a prominent variation is observed with respect to $(NH_4)_2SO_4$ concentration in the medium. More single-labeled alanine is observed in 40 mM $(NH_4)_2SO_4$ provided in the medium. In contrast, single-labeled alanine percentage is reduced when 100 mM $(NH_4)_2SO_4$ is provided in the medium, most likely due to the inhibition of the serine deamination.

10.3. Supplemental tables S1-S5

Table S 1: List of components used in DO-Ser-Gly mix.
Note: Glycine and serine are excluded from the mix.

S1 Table: Composition of Drop-Out mix used in SC medium			
S. No.	**Chemical component**	**mg/L**	**SC-Ser-Gly**
1	Adenine	18	+
2	p-Aminobenzoic acid	8	+
3	Leucine	380	+
4	Alanine	76	+
5	Arginine	76	+
6	Asparagine	76	+
7	Aspartic acid	76	+
8	Cysteine	76	+
9	Glutamic acid	76	+
10	Glutamine	76	+
11	Glycine	76	-
12	Histidine	76	+
13	Myo-inositol	76	+
14	Isoleucine	76	+
15	Leucine	76	+
16	Lysine	76	+
17	Methionine	76	+
18	Phenylalanine	76	+
19	Proline	76	+
20	Serine	76	-
21	Threonine	76	+
22	Tryptophan	76	+
23	Tyrosine	76	+
24	Uracil	76	+
25	Valine	76	+

Table S 2: List of the primers used for screening and sequencing

Other primers used for PCRs and cloning were directly indicated on the respective plasmid maps S1-S8.

S. No	Name	Sequence	Purpose	Target
1	**ASK-FN2**	CGGCCTTTTTACGGTTCCTG	Seq. sgOligo insertion	Sequencing
2	**3569**	**gactttt**GTTGTCGCAAAATAGCTATG	sgRNA	AGX1_pWS_f
3	**3570**	**aaac**CATAGCTATTTTGCGACAAC**aa**		AGX1_pWS_r
4	**3571**	gactttGGTCTCAAACTACATTGTGA		Gly1_pWS_f
5	**3572**	aaacTCACAATGTAGTTTGAGACCaa		Gly1_pWS_r
6	**3573**	gactttCCGGTGACTAAATAACCGGC		SER1_pWS_f
7	**3574**	aaacGCCGGTTATTTAGTCACCGGaa		SER1_pWS_r
8	**3634**	TGTCTCCGGTTAGTGTGTGC	Knockout screening	AGX1_f
9	**3635**	GCCCCAGAAGTGCTTTTTGG		AGX1_r
10	**3636**	TACACACGGCCCCAAATTGT		GLY1_f
11	**3637**	CACCGATCGAGGAAGCCTTT		GLY1_r
12	**3638**	AAAAAGCGTTTCGTGGAGCA		SER1_f
13	**3639**	GCTCCCTCTTGCACTTCACT		SER1_r
14	**3632**	ATCAACAACAGAGGACATATGCC	sgRNA expression cassette amplificaiton	pWS082_f
15	**3633**	ATCCTGCACTCATCTACTACCC		pWS082_r
16	**4415**	gaagtgccattccgcctgacct	TWIST adapter primers for donor DNA amplificaiton	Donor DNA amplificaiton
17	**4416**	cactgagcctccacctagcct		
18	**4619**	CTACATACATGTACATATATTTAAACATGTAAACCCGTCCATTATATTGCGACATGGAGGCCCAGAATACCCTC	pFM0340_assembly	Plasmid assembly
19	**4620**	TGAATAACTGCTCCCGTTGTAAAAGTATCCGTAGTGTGACTGAAACCTTACAGTATAGCGACCAGCATTCACATACGATTGACG	pFM0340_assembly	
20	**4656**	AGTGACCACCATCTGGCAAA	pCfB2904 SHM1 seq R2	Sequencing donor pathway modules
21	**4657**	ACATCTGCATAATAGGCATTTGCA	pCfB2904 SHM1 seq F3	
22	**4658**	AAACACTTCCTTTTTCTGGCCC	pCfB2904 SHM1 seq R3	
23	**4659**	CGTTTGCGCATCCTCTTTTT	pCfB3042 SHM2 seq F1	
24	**4660**	AGACGAAGCTTGTGTGTGGG	pCfB3042 seq R1	
25	**4661**	GCGCGATGGTGCTAAGTTAC	pCfB3042 seq F2	
26	**4662**	GCGGCGATCAAACCAGAAAT	pCfB3042 seq R2	
27	**4663**	CCAAAGGTTCTTGTTGCTGGT	pCfB3042 seq F3	
28	**4664**	CGAATTCGCGGGGTGTACAT	pCfB3042 seq R3	
29	**6643**	cccaaagctaagagtcccattttattc	pCFB3035 donor X-4 Fw	pCfB donor amplificaiton
30	**6644**	ctggtgaggatttacggtatgatcatg	pCFB3035 donor X-4 Rev	
31	**3921**	GGCCGCGCTGAGGTCTTAAT	Easy_clone_donor_fw	
32	**3922**	GGCCGCGCTGAGGGTTTAAT	Easy_clone_donor_rev	
33	**4401**	gccagggttttcccagtcacgacgttgtgtttcacaaccagagattcagg	EcGlyA expression cassette	Fw primer
34	**4402**	gccagggttttcccagtcacgacgttgTACGCAGTCTTCGGGTAGTAAAATAG	EcGlyA expression cassette	Rev primer
35	**4048**	GGACGCATGGTTATTGGTGC	pFM_0274_f	Plasmid assembly
36	**4049**	GATCGACCTTCCATGGGGTC	pFM_0274_r	
37	**4400**	gcgtgcgatgagcgacctcatgctatataaggtttcagtcacactacgg	pVB07 assembly	
38	**4401**	gccagggttttcccagtcacgacgttgtgtttcacaaccagagattcagg	pVB07 assembly	
39	**6938**	gactttGCCAAGGGTCCATACAAGGG	sgRNA	sgGDH1
40	**6939**	aaacCCCTTGTATGGACCCTTGGCaa		
41	**6940**	gactttACACATAGACCACCCTTGTA		
42	**6941**	aaacTACAAGGGTGGTCTATGTGTaa		

Table S 3: One-way ANOVA and multiple comparison of ALE growth rates

ANOVA results								
Table Analyzed	Col: One-way ANOVA							
Repeated measures ANOVA summary								
Assume sphericity?	No							
F	141.4							
P value	<0.0001							
P value summary	****							
Statistically significant (P < 0.05)?	Yes							
Geisser-Greenhouse's epsilon	0.5555							
R squared	0.9402							
Was the matching effective?								
F	1.27							
P value	0.2865							
P value summary	ns							
Is there significant matching (P < 0.05)?	No							
R squared	0.01864							
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value			
Treatment (between columns)	0.007107	4	0.001777	F (2.222, 20.00) = 141.4	P<0.0001			
Individual (between rows)	0.0001436	9	0.00001595	F (9, 36) = 1.270	P=0.2865			
Residual (random)	0.0004523	36	0.00001257					
Total	0.007703	49						
Data summary								
Number of treatments (columns)	5							
Number of subjects (rows)	10							
Number of missing values	0							
Multiple comparision								
Number of families	1							
Number of comparisons per family	10							
Alpha	0.05							
Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value			
Ev03 vs. Ev06	-0.01255	-0.01573 to -0.009365	Yes	****	<0.0001	A-B		
Ev03 vs. Ev12	-0.01936	-0.02506 to -0.01366	Yes	****	<0.0001	A-C		
Ev03 vs. Ev16	-0.02978	-0.03474 to -0.02482	Yes	****	<0.0001	A-D		
Ev03 vs. Ev21	-0.03285	-0.03450 to -0.03120	Yes	****	<0.0001	A-E		
Ev06 vs. Ev12	-0.00681	-0.01312 to -0.0005025	Yes	*	0.0338	B-C		
Ev06 vs. Ev16	-0.01723	-0.02462 to -0.009836	Yes	***	0.0002	B-D		
Ev06 vs. Ev21	-0.0203	-0.02446 to -0.01614	Yes	****	<0.0001	B-E		
Ev12 vs. Ev16	-0.01042	-0.01626 to -0.004579	Yes	**	0.0014	C-D		
Ev12 vs. Ev21	-0.01349	-0.01998 to -0.006999	Yes	***	0.0005	C-E		
Ev16 vs. Ev21	-0.00307	-0.008182 to 0.002042	No	ns	0.3296	D-E		
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
Ev03 vs. Ev06	0.0041	0.01665	-0.01255	0.0009472	10	10	18.74	9
Ev03 vs. Ev12	0.0041	0.02346	-0.01936	0.001695	10	10	16.15	9
Ev03 vs. Ev16	0.0041	0.03388	-0.02978	0.001476	10	10	28.52	9
Ev03 vs. Ev21	0.0041	0.03695	-0.03285	0.000492	10	10	94.43	9
Ev06 vs. Ev12	0.01665	0.02346	-0.00681	0.001876	10	10	5.134	9
Ev06 vs. Ev16	0.01665	0.03388	-0.01723	0.002199	10	10	11.08	9
Ev06 vs. Ev21	0.01665	0.03695	-0.0203	0.001236	10	10	23.23	9
Ev12 vs. Ev16	0.02346	0.03388	-0.01042	0.001737	10	10	8.483	9
Ev12 vs. Ev21	0.02346	0.03695	-0.01349	0.00193	10	10	9.883	9
Ev16 vs. Ev21	0.03388	0.03695	-0.00307	0.00152	10	10	2.856	9

Table S 4: One-way ANOVA and multiple comparison of formate gradient

ANOVA results								
Table Analyzed	Data 9							
Repeated measures ANOVA summary								
Assume sphericity?	No							
F	99.47							
P value	<0.0001							
P value summary	****							
Statistically significant (P < 0.05)?	Yes							
Geisser-Greenhouse's epsilon	0.1743							
R squared	0.9521							
Was the matching effective?								
F	1.593							
P value	0.1734							
P value summary	ns							
Is there significant matching (P < 0.05)?	No							
R squared	0.005416							
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value			
Treatment (between columns)	0.02201	14	0.001572	F (2.440, 12.20) = 99.47	P<0.0001			
Individual (between rows)	0.0001259	5	0.00002517	F (5, 70) = 1.593	P=0.1734			
Residual (random)	0.001106	70	0.0000158					
Total	0.02324	89						
Data summary								
Number of treatments (columns)	15							
Number of subjects (rows)	6							
Number of missing values	0							
Multiple comparision								
Number of families	1							
Number of comparisons per family	14							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value	G-?		
75 vs. 0	0.04188	0.03870 to 0.04506	Yes	****	<0.0001	A	0	
75 vs. 2	0.04188	0.03870 to 0.04506	Yes	****	<0.0001	B	2	
75 vs. 5	0.03438	0.01896 to 0.04980	Yes	**	0.0015	C	5	
75 vs. 10	0.01557	0.003132 to 0.02801	Yes	*	0.0203	D	10	
75 vs. 25	0.003611	0.0007384 to 0.006483	Yes	*	0.0199	E	25	
75 vs. 50	0.001596	0.0001816 to 0.003010	Yes	*	0.0311	F	50	
75 vs. 100	0.0003873	-0.0006478 to 0.001422	No	ns	0.658	H	100	
75 vs. 125	0.001276	-0.003556 to 0.006108	No	ns	0.8905	I	125	
75 vs. 200	0.003043	0.001372 to 0.004715	Yes	**	0.0039	J	200	
75 vs. 250	0.004662	0.001435 to 0.007888	Yes	*	0.011	K	250	
75 vs. 300	0.004784	0.002401 to 0.007167	Yes	**	0.0025	L	300	
75 vs. 500	0.01104	0.004098 to 0.01798	Yes	**	0.0072	M	500	
75 vs. 750	0.01775	0.008739 to 0.02676	Yes	**	0.0027	N	750	
75 vs. 1000	0.03813	0.02729 to 0.04897	Yes	***	0.0002	O	1000	
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
75 vs. 0	0.04188	0	0.04188	0.000752	6	6	55.69	5
75 vs. 2	0.04188	0	0.04188	0.000752	6	6	55.69	5
75 vs. 5	0.04188	0.007497	0.03438	0.003649	6	6	9.421	5
75 vs. 10	0.04188	0.02631	0.01557	0.002944	6	6	5.289	5
75 vs. 25	0.04188	0.03827	0.003611	0.0006797	6	6	5.312	5
75 vs. 50	0.04188	0.04028	0.001596	0.0003347	6	6	4.768	5
75 vs. 100	0.04188	0.04149	0.0003873	0.000245	6	6	1.581	5
75 vs. 125	0.04188	0.0406	0.001276	0.001144	6	6	1.116	5
75 vs. 200	0.04188	0.03884	0.003043	0.0003956	6	6	7.693	5
75 vs. 250	0.04188	0.03722	0.004662	0.0007636	6	6	6.104	5
75 vs. 300	0.04188	0.03709	0.004784	0.000564	6	6	8.484	5
75 vs. 500	0.04188	0.03084	0.01104	0.001642	6	6	6.721	5
75 vs. 750	0.04188	0.02413	0.01775	0.002133	6	6	8.323	5
75 vs. 1000	0.04188	0.00375	0.03813	0.002566	6	6	14.86	5

Table S 5: One-way ANOVA and multiple comparison for the gradient of ammonium sulfate

ANOVA results								
Table Analyzed	Data 10							
Data sets analyzed	A-G							
ANOVA summary								
F	8.932							
P value	0.0004							
P value summary	***							
Significant diff. among means (P < 0.05)?	Yes							
R squared	0.7929							
Brown-Forsythe test								
F (DFn, DFd)	0.3302 (6, 14)							
P value	0.9099							
P value summary	ns							
Are SDs significantly different (P < 0.05)?	No							
Bartlett's test								
Bartlett's statistic (corrected)								
P value								
P value summary								
Are SDs significantly different (P < 0.05)?								
ANOVA table	SS	DF		MS	F (DFn, DFd)	P value		
Treatment (between columns)	0.0001241	6		0.00002068	F (6, 14) = 8.932	P=0.0004		
Residual (within columns)	0.00003241	14		0.000002315				
Total	0.0001565	20						
Data summary								
Number of treatments (columns)	7							
Number of values (total)	21							
Multiple comparision								
Number of families	1							
Number of comparisons per family	6							
Alpha	0.05							
Dunnett's multiple comparisons test	Mean Diff.	95.00% CI of diff.		Below threshold?	Summary	Adjusted P Value	C-?	
40 mM vs. 10 mM	0.007067	0.003448 to 0.01069		Yes	***	0.0003	A	0.25x
40 mM vs. 20 mM	0.001133	-0.002485 to 0.004752		No	ns	0.8656	B	0.5x
40 mM vs. 60 mM	0.0008667	-0.002752 to 0.004485		No	ns	0.9524	D	1.5x
40 mM vs. 80 mM	-0.0009333	-0.004552 to 0.002685		No	ns	0.9352	E	2x
40 mM vs. 100 mM	0.0007333	-0.002885 to 0.004352		No	ns	0.9773	F	2.5x
40 mM vs. 120 mM	0.0026	-0.001018 to 0.006218		No	ns	0.2104	G	3x
Test details	Mean 1	Mean 2		Mean Diff.	SE of diff.	n1	n2	q	DF
40 mM vs. 10 mM	0.0452	0.03813		0.007067	0.001242	3	3	5.688	14
40 mM vs. 20 mM	0.0452	0.04407		0.001133	0.001242	3	3	0.9122	14
40 mM vs. 60 mM	0.0452	0.04433		0.0008667	0.001242	3	3	0.6976	14
40 mM vs. 80 mM	0.0452	0.04613		-0.0009333	0.001242	3	3	0.7513	14
40 mM vs. 100 mM	0.0452	0.04447		0.0007333	0.001242	3	3	0.5903	14
40 mM vs. 120 mM	0.0452	0.0426		0.0026	0.001242	3	3	2.093	14

Table S 6: One-way ANOVA and multiple comparison for high CO_2 dependency of the RGP

ANOVA results								
Table Analyzed	Data 12							
Repeated measures ANOVA summary								
Assume sphericity?	No							
F	969.7							
P value	<0.0001							
P value summary	****							
Statistically significant (P < 0.05)?	Yes							
Geisser-Greenhouse's epsilon	0.4891							
R squared	0.9979							
Was the matching effective?								
F	0.7395							
P value	0.5163							
P value summary	ns							
Is there significant matching (P < 0.05)?	No							
R squared	0.0005071							
ANOVA table	SS	DF	MS	F (DFn, DFd)	P value			
Treatment (between columns)	0.003598	3	0.001199	F (1.467, 2.934) = 969.7	P<0.0001			
Individual (between rows)	0.000001829	2	9.146E-07	F (2, 6) = 0.7395	P=0.5163			
Residual (random)	0.000007421	6	0.000001237					
Total	0.003607	11						
Data summary								
Number of treatments (columns)	4							
Number of subjects (rows)	3							
Number of missing values	0							
Multiple comparision								
Number of families	1							
Number of comparisons per family	6							
Alpha	0.05							
Tukey's multiple comparisons test	Mean Diff.	95.00% CI of diff.	Below threshold?	Summary	Adjusted P Value			
1 vs. 2.5	-0.01254	-0.01914 to -0.005941	Yes	*	0.0144	A-B		
1 vs. 5	-0.03894	-0.04199 to -0.03588	Yes	****	<0.0001	A-C		
1 vs. 10	-0.04052	-0.04512 to -0.03592	Yes	****	<0.0001	A-D		
2.5 vs. 5	-0.0264	-0.03592 to -0.01687	Yes	**	0.007	B-C		
2.5 vs. 10	-0.02798	-0.03437 to -0.02159	Yes	**	0.0016	B-D		
5 vs. 10	-0.001582	-0.007232 to 0.004067	No	ns	0.4219	C-D		
Test details	Mean 1	Mean 2	Mean Diff.	SE of diff.	n1	n2	q	DF
1 vs. 2.5	0	0.01254	-0.01254	0.0009522	3	3	18.62	2
1 vs. 5	0	0.03894	-0.03894	0.0004414	3	3	124.8	2
1 vs. 10	0	0.04052	-0.04052	0.0006634	3	3	86.37	2
2.5 vs. 5	0.01254	0.03894	-0.0264	0.001375	3	3	27.15	2
2.5 vs. 10	0.01254	0.04052	-0.02798	0.0009222	3	3	42.91	2
5 vs. 10	0.03894	0.04052	-0.001582	0.0008154	3	3	2.744	2

11. Acknowledgements

This work would not have been possible without the help and support of many different people. Therefore, I would like to use this opportunity to express my gratitude.

First and foremost, I would like to thank (תודה) **Dr. Arren Bar-Even** for giving me an opportunity to join his group as a Ph.D. student and to work on engineering yeast metabolism. Arren's supervision and his brilliant ideas on metabolic systems influenced me a lot while working on my Ph.D. project. The untimely demise of Arren is an immeasurable loss and devastation.

I want to thank (danke) **Dr. Fabian Machens**, who supervised me in working on this yeast project. After the demise of Arren, Fabian took greater responsibility for helping me with day-to-day requirements. Also, Fabian helped me read my reports and thesis partly.

I would like to especially thank **Prof. Mark Stitt** for being my mentor, PAC chair, and official supervisor, particularly after the demise of Arren. Mark has been in many roles in my Ph.D.; I heartfully thank Mark for taking the time to read my thesis and for the many roles he has taken to help me in my Ph.D. time at MPIMP.

I would like to sincerely thank **Prof. Bernd Müller-Röber** for his valuable suggestions as a PAC member and for being my second supervisor after the demise of Arren. I would like to thank **Dr. Friedrich Kragler** for being my PAC member and for giving me valuable inputs. I would like to especially thank Dr. Friedrich Kragler for reading my reports and giving me suggestions to improve my writing.

Special thanks to **Prof. Lothar Willmitzer** for taking charge of the group now and helping me with the necessary extension in order to stay at the MPIMP and finish my Ph.D.

Special thanks to **Dr. Ina Talke** for helping me with the whole Ph.D. process and guiding me on the roles to finish the responsibilities as a Ph.D. student.

I want to thank MPIMP for hosting and supporting me as a doctoral researcher. I want to thank the each and every person in the group of **AG Bar-Even** for their support during my Ph.D.

I want to thank my students, **Ayesha** and **Abdel**, for their help during my Ph.D. project progression. Special thanks to Ayesha for supporting me for a long time during the project.

Thanks to **Dr. Arun Sampath Kumar** for helping with microscopy. Thanks to **Änne Michaelis** for LC-MS support. A special thanks to each and every person of central teams like IT, media kitchen and house technique for their support.

Thanks to **Dr. Nico J Claassens** for his input on some experiments in my thesis.

Special thanks (நன்றி) to **Dr. Vijay Jayaraman** for helping me to understand mutations in protein structures.

Most importantly, thanks to my parents (**Subhalakshumma** and **Jayarami Reddy**) for raising me as a decent kid and supporting my education up to my M.Sc., which allowed me to work

to gain research experience and further join for my Ph.D. at MPIMP. I would like to thank my wife **Supraja**, my son **Abhayanth Reddy**, and my daughter **Anika** for their loving support. Special thanks to my wife, who supported me so much that more than 95% of the work that I presented in this thesis was generated after my wife and son joined me in the Potsdam on 3rd March 2020 (Before COVID lockdowns). I want to thank my uncles **Sreenivasula Reddy** (who passed away during my Ph.D.) and **Venkata Subba Reddy** for their support in my education. Special thanks to my grandparents for their best wishes, who also passed away during my Ph.D. time.

Special thanks (நன்றி) to my friend **Yoga** for reading my thesis completely as a third person in crucial times. Thanks (Спасибо) to my friend **Katrina** for her moral support. Thanks to my other friends who helped my family and me during the difficult times.

ధన్యవాదములు/நன்றி /धन्यवाद/Dhanyawad — నమస్కారము/நன்றி/नमस्ते/Namaste

12. Author contribution

Viswanada Reddy Bysani Kondagari (V.R.B.K.): All the experiments and analyses were performed by V.R.B. under the supervision of Arren Bar-Even and Fabian Machens. I am also involved in conceptualizing *EcSHMT* engineering, FL100 engineering, reverse engineering of mutation, and pyruvate synthesis and energy module engineering to establish growth on serine formate with glucose and also involved in supervision, visualization, Investigation, data curation, writing & editing.

Ayesha Syed Mamoor Alam (A.S.M.A.): Involved in screening serine biosensor strains, i.e., VBS08, VBS09, and VBS10 strains with SHMTox. She was also involved in constructing FL100 serine biosensor strains and labeling data analysis under **V.R.B.K.**'s supervision.

Abdelrahman Alnahhal (A.A.): Involved in screening VBS20 strains for the pathway module under the supervision of **V.R.B.K.**

Fabian Machen (F.M.): Supervised **V.R.B.K.**, especially after the demised **A.B.E.**, involved in reading part this thesis (materials and methods).

Arren Bar-Even (A.B.E.): Conceptualized the complete research, supervised (until his demise), and acquired funding to conduct this research.

Prof. Mark Stitt (M. S.): Supervision and was involved in reading and editing the complete thesis after the demised of Arren Bar-Even as an official supervisor for **V.R.B.K.**'s thesis.

Author declaration/conflict of interest

A.B.E. was co-founder of b.fab company, which aimed to commercialize engineered C1-assimilation microorganisms. The company was not involved in any way in conducting, funding, or influencing the research.

Funding

This project was funded by the Max-Planck-Gesellschaft via the Max Planck Institute of Molecular Plant Physiology.

ORCID

Viswanada Reddy Bysani Kondagari: 0000-0002-1762-8457